Molecular Dynamics Studies of the Folding of α-Helical Cross-Linked Peptides

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät der

Universität Zürich

von

Beatrice Paoli

aus

Italien

Promotionskomitee

Prof. Dr. Amedeo Caflisch (Vorsitz)

Prof. Dr. Peter Hamm

Zürich 2009

Summary

The helix-coil transition represents the simplest scenario in protein folding, but the details of its kinetics are not fully understood. Because the helix is a common secondary structural motif in proteins, it is important to understand the folding process of the α-helix in detail as a step to understand the protein folding mechanism. A well-established way of studying physicochemical properties of an α-helix is to use model peptides with an enhanced helix-propensity. In the last years, the employment of peptides with an attached photoswitchable cross-linker, whose photo-isomerization can induce the unfolded or folded state, has been established as an efficient experimental tool to control the helix stability. The observation, by means of time resolved infrared spectroscopy, of a non-exponential, stretched kinetics of folding at low temperature suggested that the system cannot be described simply by a two-state model.

In this thesis, molecular dynamics simulations studies at atomic resolution level were performed to shed light on the physical picture which has started to emerge from infrared experiments. The investigation of the site-specific folding kinetics from both experimental and computational sides revealed that the folding rates differ from site to site suggesting that the folding of a constrained α-helix is noncooperative in contrast with conventional nucleation-propagation models according to which one could expect similar rates. The Kinetic Grouping Analysis method unmasked the existence of discrete traps

(i.e., non-native free energy basins) along parallel folding pathways, which render the overall kinetics non-exponential providing evidence that the rate limiting step in the formation of a constrained α-helix is the escape from these traps, rather than the nucleation rate.

To investigate the effect of the sequence on the folding process three cross-linked peptides with different amount of bulky side chains were investigated. Again the folding kinetics at low temperature are stretched exponentials but different sequences show different folding rates indicating that the primary structure has a strong influence on the free-energy surface.

Finally, cut based free energy profiles revealed that the helical ensemble (i.e., native) of the cross-linked peptides is divided by a free energy barrier into two regions having different orientations of the side chains (especially Arg10) with respect to the cross-linker. Moreover, the transition from one orientation to the other is hindered by the presence of the bulky cross-linker. To investigate this effect, the cross-linker was replaced with a distance constraint mimicking the effect of a "non-bulky" cross-linker and eventually removed it. When the space previously occupied by the cross-linker is made free either by replacing or removing the cross-linker, the folding kinetics becomes faster and simpler compared to the cross-linked case. This indicates that the main effect of the cross-linker on the overall kinetics is due to its entanglement with the side chains rather than its effect as a distance constraint.

Zusammenfassung

Der Helix-Coil Übergang stellt das einfachste Szenario in der Proteinfaltung dar, aber die Details, insbesondere deren Kinetik, wurden noch nicht ganz verstanden. Da die Helix ein häufiges sekundärstruktur Motiv in Proteinen ist, ist es wichtig, den Faltungsprozess der α-Helix im Detail zu verstehen, um den Proteinfaltungsmechanismus als Ganzes zu verstehen. Eine etabilierte Prozedur, um die physikalisch-chemischen Eigenschaften einer α-Helix zu untersuchen, ist die Verwendung von Peptiden mit einer verstärkten Helixtendenz. In den letzten Jahren hat sich die Verwendung von Peptiden mit einem angehängten photoschaltbaren Cross-Linker, deren Photoisomerisierung den gefalteten und den entfalteten Zustand induzieren kann, als ein effizientes Hilfsmittel zur experimentellen Kontrolle der Helixstabilität durchgesetzt. Die Beobachtung einer nicht-exponentiellen gestreckten Faltungskinetik durch zeitaufgelöste Infrarotspektroskopie bei tiefen Temperaturen hat Indizien geliefert, dass das System nicht durch ein einfaches zwei-Zustände Modell beschrieben werden kann.

In dieser Doktorarbeit wurden Moleküldynamiksimulationen auf atomar aufgelöstem Niveau durchgeführt, um Aufschluss über das physikalische Bild zu geben, das seinen Ursprung in den Infrarot-Experimenten hatte. Die Untersuchung der ortspezifischen Faltungskinetik mit experimentellen und rechnerischen Mitteln hat gezeigt, dass sich die Faltungsraten von Ort zu Ort unterscheiden, was darauf hinweist, dass die Faltung einer eingeschränkten

α-Helix unkooperativ ist. Im Gegensatz dazu können in konventionellen Nukleationsmodellen ähnliche Raten erwartet werden. Die Kinetic Grouping Analysis Methode deckt die Existenz von ungefalteten Minima der freien-Energiefläche entlang parallelen Faltungswegen auf, welche die allgemeine Kinetik nicht-exponentiell machen. Dies bedeutet, dass der geschwindigkeitlimitierende Schritt in der Bildung einer eingeschränkten α-Helix nicht die Nukleationsrate ist, sondern das Verlassen der Minima.

Um den Effekt der Sequenz auf den Faltungsprozess zu erforschen, wurden drei cross-gelinkte Peptide mit unterschiedlicher Anzahl von sperrigen Seitenketten untersucht. Nochmals ist die Faltungskinetik bei tiefen Temperaturen eine gestreckte Exponentialfunktion, aber unterschiedliche Sequenzen zeigen unterschiedliche Faltungsraten und weisen darauf hin, dass die primäre Struktur einen starken Einfluss auf die freien-Energiefläche hat.

Cut-based-Free-Energy-Profiles ergaben schliesslich, dass das helixförmige Ensemble des cross-gelinkten Peptides durch eine Barriere in der freien Energie in zwei Regionen mit unterschiedlichen Orientierungen der Seitenketten (insbesondere Arg10) im Bezug auf den Cross-Linker getrennt wird. Zudem wird der Übergang von einer Orientierung zur anderen von der Anwesenheit des sperrigen Cross-Linkers behindert. Um diesen Effekt zu untersuchen, wurde der Cross-Linker durch eine Abstandsbedingung ersetzt, die den Effekt eines nicht-sperrigen Cross-Linkers imitiert, und wurde schliesslich ganz entfernt. Wenn der Raum, der vorher vom Cross-Linker besetzt war, entweder durch den Ersatz oder die Entfernung des Cross-Linkers befreit wird, wird die Faltungskineitk im Vergleich mit dem cross-gelinkten Fall schneller und einfacher. Dies zeigt, dass der wichtigste Effekt des Cross-Linkers auf die allgemeine Kinetik eheran dessen Verschränkung mit den Seitenketten als an dessen Effekt als Abstandsbedingung liegt.

List of publications

α-helix folding in the presence of structural constraints.

J.A. Ihalainen[‡], B. Paoli[‡], S. Muff, E.H.G Backus, J. Bredenbeck, G.A. Woolley, A. Caflisch, and P. Hamm

[‡] These two authors contributed equally to this work

[*PNAS*, **2008**, *105(28)*, 9588-9593]

Bulky side chains and non-native salt bridges slow down the folding of a cross-linked helical peptide: A combined molecular dynamics and time-resolved infrared spectroscopy study.

B. Paoli, M. Seeber, E.H.G Backus, J.A. Ihalainen, P. Hamm, and A. Caflisch

[*JPC B*, **2009**, *113(13)*, 4435-4442]

Cross-linked helical peptides have complex free-energy surface

B. Paoli et al.

[*To be submitted*]

Contents

Chapter 1

Introduction

1.1 Protein Folding

Proteins are polypeptides which play a central role in biology due to their functional properties. A great challenge in the field of protein dynamics is to understand how proteins acquire their three-dimensional structure, the prerequisite to fulfill their function. This question has gained additional significance because of the growing interest in diseases that result from protein unfolding, misfolding, and aggregation, which range from Creutzfeld-Jacob disease to cancer. The experiment on ribonuclease performed by Anfinsen [1] in early 1960s demonstrated that proteins can refold spontaneously into their native structure after denaturation. This means that the native state of a protein is entirely determined by its amino acids sequence and that the native state of a protein is the global minimum of the Gibbs free-energy surface. This is often referred to as the *thermodynamic hypothesis* of protein folding. However, the sequence of a protein must meet a second requirement: it must fold in a biologically relevant time. This can be stated as the *kinetic hypoth-*

esis of protein folding. Given the astronomically large number of possible conformations for a polypeptide chain, the question arises on how a protein can attain its native conformation in a reasonable amount of time. For example one can estimate that given a protein of 124 amino acids (like the ribonuclease studied by Anfinsen), if each of them can assume six different discrete backbone conformations, there are $6^{124} \simeq 10^{96}$ states available. If the protein should find its native state by random search (Fig. 1.1a), this could take longer than the age of the universe, whereas it is known from experiments that polypeptide chains fold in the time frame of microseconds to minutes. This problem is known as *Levinthal's paradox* [2]. Levinthal solved the paradox adducing that proteins fold following a "folding pathway", rather than a random search. Therefore, only a small part of the conformational space needs to be explored, explaining the experimental findings of fast folding. The pathway perspective implies that folding proceeds from a specific point A to N (Fig. 1.1b). But protein folding does not involve starting from one specific conformation. The denatured state of a protein is not a single point of the energy landscape but it is all the points except N. In this sense, the solution proposed by Levinthal is a paradox itself: if the denatured state is random, there cannot be a unique pathway. A widely accepted solution to Levinthal's paradox has been provided by the so-called "new view" of protein folding [3–7] developed in analogy to spin glass theory and polymer physics [8–10] and based on the statistical characterization of the energy landscape. According to this new view, the folding ensemble moves on a funnel shaped free-energy landscape towards the native state, which is the global free-energy minimum as suggested by Anfinsen. The new view recognizes that the solution to Levinthal's paradox is: "funnels, not tunnels" [11]. According to the "new view", the folding process does not involve a series of

mandatory steps (a well-defined pathway) between partly folded states (like in simple chemical reactions), but rather a stochastic search of the many conformations accessible to a polypeptide chain. What is happening is that each denatured conformation follows its own trajectory to reach the global minimum (satisfying Anfinsen's thermodynamic hypothesis) by many different routes. Additionally, they do it in a direct and rapid way (satisfying Levinthal's concern). A smooth energy landscape allows rapid transitions (Fig. 1.1c), whereas a rugged energy surface has kinetic traps which slow

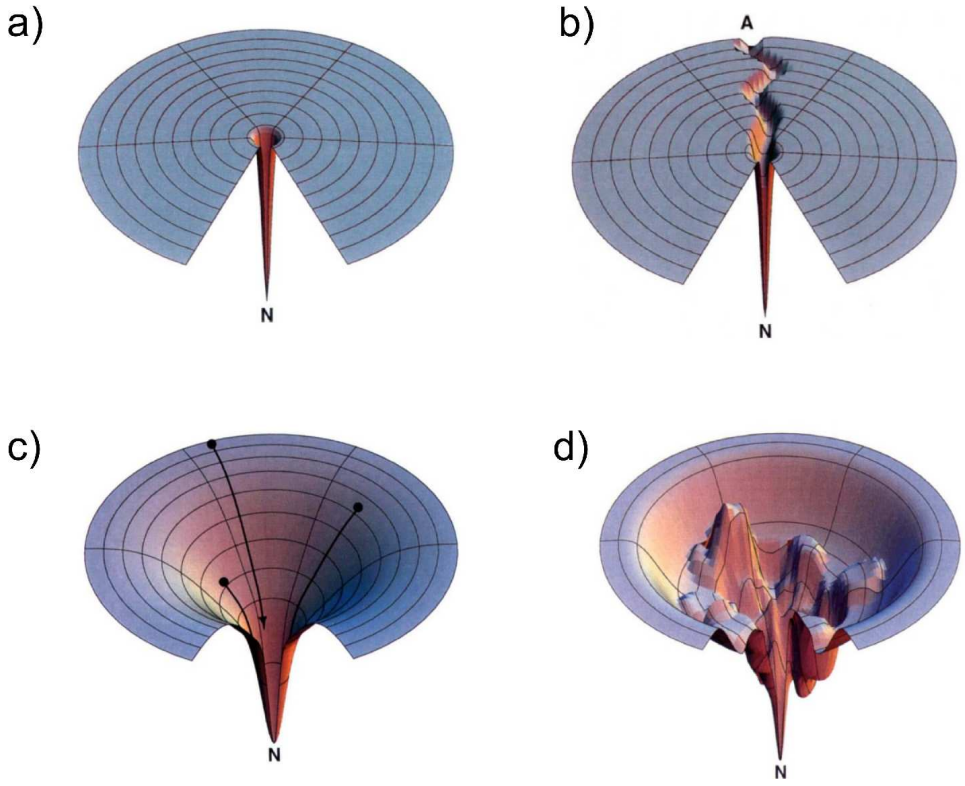


Figure 1.1: Energy landscape perspective. a) *Flat golf course*: the energy landscape as it would be if Levinthal's paradox were a problem. b) "Folding pathway" as suggested by Levinthal: the folding proceeds from a specific point A to the point N. c) Smooth energy landscape allowing fast transitions. d) A rugged energy landscape, with some kinetic traps which slow down folding. Pictures are taken from reference [3] and somewhat modified.

down folding (Fig. 1.1d) and occurs in problems in which there are many competing interactions in the energy function. This competition is called *frustration*.

Different models have been proposed to explain the high speed of protein folding. The *hierarchical model* [12, 13], proposes that protein folding starts with the formation of elements of secondary structure already in the unfolded state regardless of the tertiary contacts. Essentially the local propensities of adjacent amino acids drive the formation of backbone-backbone hydrogen bonds which are responsible for the construction of secondary structure. This step of local arrangement can be very fast depending on the strength of the local preferences of the amino acids. These elements of secondary structure assemble into the tightly packed native tertiary structure by means of a *diffusion-collision* mechanism (Fig.1.2a). The *hydrophobic-collapse model* [15, 16] (Fig.1.2b) suggests that the protein buries its hydrophobic side chains from the solvent forming a collapsed intermediate also known as molten globule. The native state develops by a search within the conformationally restricted area of the molten globule. In the *nucleation-condensation* mechanism [17, 18] a folding nucleus involving a low number of close residues, catalyzes further folding in a stepwise manner (Fig.1.2c). The primary nucleus mainly consist of a few adjacent residues which have some correct secondary structure interactions. The last two models are compatible with the theory of funneled energy landscape [3, 6]: the nucleus is the rate-limiting step of the folding reaction after which the growth of favorable native interactions dramatically decreases the enthalpy of the protein compensating the entropy loss and eventually overtaking it.

The great advances in the fields of protein folding during the last 20 years are due in large part to technological advances and cooperation between

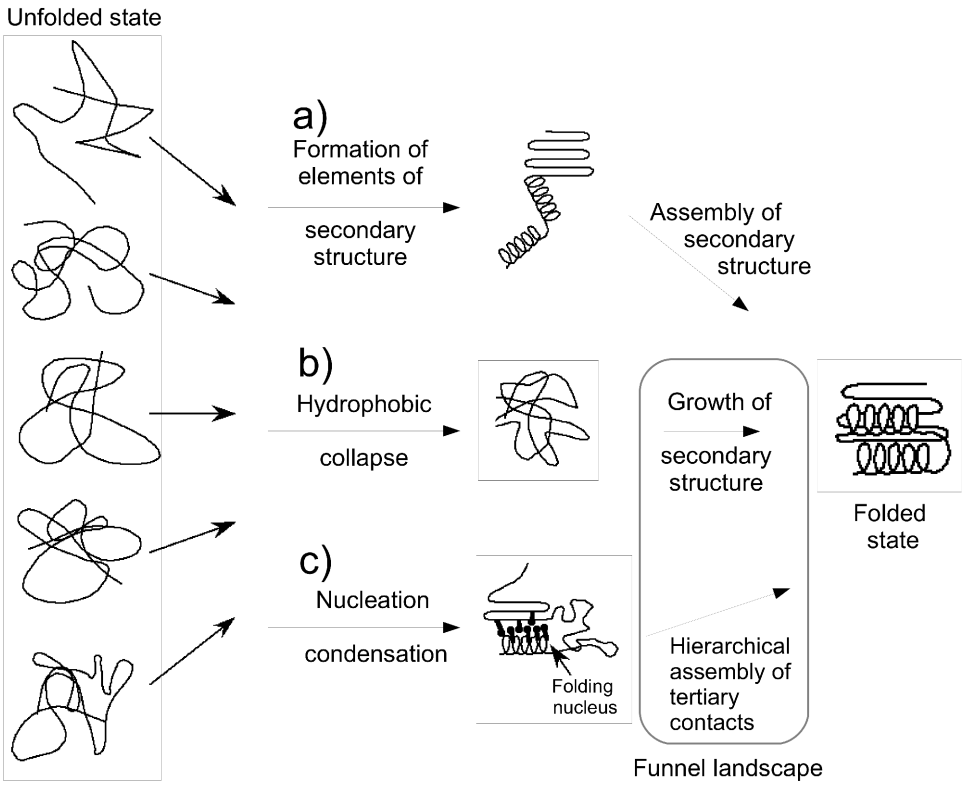


Figure 1.2: Possible mechanisms for protein folding. a) *Hierarchical model.* Protein folding starts with the formation of elements of secondary structure independently from tertiary structure. These elements assemble into the tightly packed tertiary structure by a diffusion-collision mechanism. b) *Hydrophobic collapse model.* The initial event is a collapse of the protein molecule, driven by hydrophobic effect. c) N*ucleation-condensation mechanism.* Early formation of a protein folding nucleus catalyzes further folding. The figure is adapted from [14].

theoreticians and experimentalists. Among the recently developed methods for studying protein folding, we have nuclear magnetic resonance (NMR) and single molecule technique. NMR experiments can provide detailed insights into the structure and dynamics of unfolded states, and studies of several proteins have been reported, including lysozyme [19–21], 434 repressor [22], FK506-binding protein [23], BPTI [24], protein G [25], barnase [26], an SH3

domain [27, 28], and staphylococcal nuclease [29, 30].

Single molecule method can be coupled with different spectroscopic techniques. Currently, the two most common single molecule methods to study protein folding are atomic force microscopy [31,32] and Förster resonance energy transfer (FRET) (for a review see reference [33] and references therein). Additionally, the study of very fast folding process, in the time-scale of microseconds or less, constitutes another important area in current protein folding studies. In fact, the most common secondary structure elements in proteins -α-helix and β-sheet- show fast folding rate. For this reason the study of the folding dynamics of monomeric helices and β-hairpins requires a fast initiation method. In particular the helix dynamics have been studied for the most part by using laser induced T-jump methods to perturb the equilibrium, and the subsequent conformational dynamics are detected either by time-resolved IR spectroscopy [34–38], by fluorescence [39, 40], or by Raman scattering techniques [41]. Temperature jumps can be in principle very fast and are only limited by the thermal equilibration time of the bulk solvent that is in the range of a few to tens of picoseconds. Nevertheless, most T-jump studies have been limited to the 10 ns regime for technical reasons. Moreover these experiments have often been modeled by "kinetic zipper" models [40, 42, 43]. Each residue of the polypeptide chain is considered to exist in one of two possible configurations, coil or helical. Although this type of analysis successfully describes a large number of helix relaxation data, molecular dynamic simulations of small peptides [44–48] demonstrated the existence of low free-energy traps in the unfolded state. Such misfolded traps are completely ignored in kinetic-zipper models. Hamm and coworkers [49] chose another approach to initiate conformational transitions in small peptides. By incorporating a photoswitch into a peptide that undergoes a

photochemical reaction after electronic excitation, one may alter the conformation of the peptide in a controlled way. The switching time is ultrafast in most cases ($\sim$ 1 ps), and therefore does not overlap with potentially interesting timescales of the peptide. The fact that is now possible to perform long molecular dynamics (MD) trajectories, at least using an implicit solvation model [50], makes these systems suitable to be studied from a computational point of view. MD simulations combined with experimental results provide an enhanced understanding of the mechanisms of protein folding. In order to model the dynamics of a physical system it is necessary to define an energy function which describes the energy of the system in a given conformation. Owing to the size of the considered phase space, the numerical treatment of the problem implies a large computational effort and in most cases requires high-performance cluster-computers. Many MD algorithms have been written in the past decades based on empirical energy functions. An example is the one used in the program CHARMM [51] which is the one used throughout this work.

The present work focuses exactly on the study of the folding of photoswitchable peptides by means of MD simulations. In the next sections the main features of photoswitchable peptides will be presented, as well as recent methods developed in our group to investigate their free-energy landscape.

1.2 Photoswitchable peptides

As already mentioned in the previous section, one of the fundamental issues for biophysics is to understand how protein fold to their native structure. To fully address this question it is necessary to explore both the energetics and the dynamics of folding. In exploring the dynamics of protein folding it is

useful to study both natural occurring proteins as well as model polypeptide systems, to understand the factors governing the secondary structure formation. Moreover, for relatively short helical peptides, folding is expected to occur in the submicrosecond time scale. Thus, to study these fast events it is necessary to have not only fast spectroscopic probes able to distinguish the helical from the random structure, but the folding process needs to be triggered on the nanosecond and even picoseconds time scale. An approach to initiate peptide folding is the synthesis of peptides incorporating a photoactive molecule [52]. This approach offers the ability to reversibly control secondary structure content. In particular, two alanines at positions 3 and 14 are replaced by cysteine to allow for the attachment of the photoswitchable cross-linker. Switching the linker between cis (i.e., closed) and trans (i.e., opened) conformations modifies the energy landscape of the peptide and allows for large changes in helicity (Fig.1.3). Three helical turns are required to form an α-helix out of the 12 residues bridged by the linker. These three turns do not fit well into the distance spanned by the linker in the cis form; thus, helicity is greatly reduced when the linker is in this conformation. Only

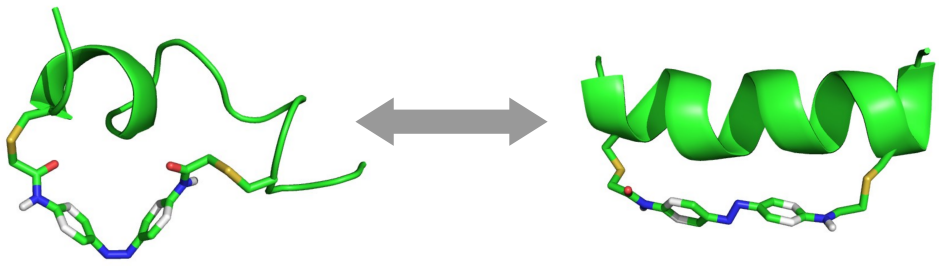


Figure 1.3: Example of photoswitchable peptide. (Top) Cross-linker in cis (left) and trans (right) conformations. (Bottom) Schematic model of cis-peptide and trans-peptide, illustrating the conformational transition induced by the cross-linker. Hydrogens and side chains are omitted for clarity.

after stretching the linker by switching it to trans the formation of the α-helix can take place. The isomerization of the linker projects the unfolded ensemble onto an energy landscape that favors the folded state, and folding can be observed as the dominant kinetic contribution.

The main experimental observation is the non-exponential kinetics of folding when the system temperature is lowered. It is known from chemical kinetics that if the system has to surmount a single barrier, the kinetics is single-exponential. In this case the classical two-state model is appropriate to describe the transition to the folded state. When the folding kinetics becomes non-exponential, the system cannot be described by a two-state model. A well known model applying in this case is the "downhill" one, according to which no free-energy barrier exists between the folded and the unfolded state [6]. Anyway, the microscopic physical mechanisms underlying the complex kinetics observed experimentally are still unclear. In this respect, MD simulations studies at atomic resolution level can provide a better understanding of the physical picture which has started to emerge from IR experiments. In this respect, we address ourselves the following questions:

1. What are the reasons for the complex kinetics of folding?

2. Can the simulations be used to derive a quantitative description of the free-energy surface?

3. Is it possible to predict mutants with different folding rates?

4. How do bulky side chains influence the kinetics of α-helix formation?

To answer these questions, we studied three cross-linked peptides (Fig.1.4) with differ amount of bulky side chains and with different salt bridges. The folding process upon ultrafast switching is emulated in the MD simulations by

sudden switch of the dihedral potential of the N=N bond in the middle point of the photoswitchable cross-linker. Two different techniques, developed in our group, were applied to shed light on the kinetics of the folding process and analyze its free-energy surface. They will be presented in the next Sections.

Sequence	Short name	Nr. bulky side chains	Salt Bridges i, i+4
-AACAR^{5}AAAAR^{10}AAACR^{15}A-	AAAAR	3	No
-EACAR^{5}EAAAR^{10}EAACR^{15}Q-	EAAAR	6	Yes
-EMCAR^{5}EMAAR^{10}EMACR^{15}Q-	EMAAR	9	Yes

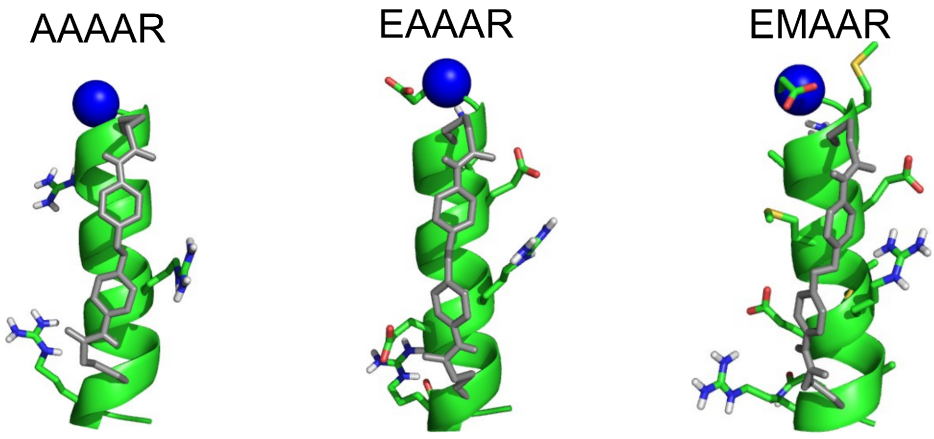


Figure 1.4: Summary of the photoswitchable peptides investigated. (Top) Summary table reporting the complete sequence and main characteristics of the cross-linked peptides: number of bulky side chains and salt bridges. (Bottom) Schematic illustration of the three peptides with their bulky side chains.

1.3 The free-energy surface of peptide and protein folding

The thermodynamics and kinetics of a system are not governed only by the energetics, i.e. the potential energy (internal energy or enthalpy), but also by the entropy that, in the case of a protein, is the conformational entropy. The major role played by the entropy contributions indicates that the analysis of the free-energy surface is more important than the analysis of the potential energy surface [53–56]. The free-energy of a protein is defined as:

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

where the enthalpy ΔH depends on van der Waals and electrostatic interactions, as well as the effects of the solvent, and the entropy ΔS accounts for the amount of available conformations and thus for its flexibility. During the folding the loss of entropy is counterbalanced by favorable interactions between protein atoms and the process is governed by a complex free-energy surface. The common way to investigate and display the free-energy surface is to study it as a function of a small number of *order parameters*, i.e., suitably chosen macroscopic quantities that distinguish the different states of the protein [57]. The main disadvantage of this commonly used projections is the possibility of hiding essential information concerning the free-energy surface [45, 46, 48, 58], in particular the conformational heterogeneity of the denatured state. Another approach to the study of the protein free-energy landscape is to use unprojected surfaces such as complex network [46] and graph analysis [45] which are able to capture the actual complexity of the protein free-energy surface. These approaches show that the picture of protein folding, obtained by projecting the free-energy onto an arbitrarily chosen

progress variable, is not consistent with the complexity of the actual free-energy surface [48].

In Fig. 1.5 is plotted the projection of the folding free-energy surface of EAAAR peptide onto the number of native hydrogen bonds (i.e., those between residues at positions i and i+4). As we will see in next sections, this picture is not sufficient to explain the complexity of the free-energy surface of this peptide.

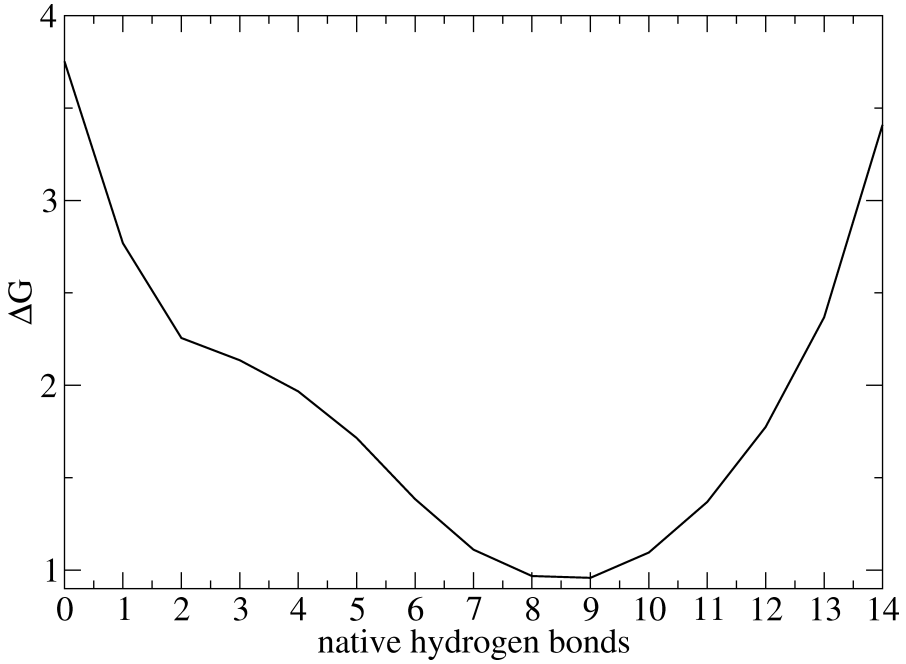


Figure 1.5: Reduction of complexity. In the case of EAAAR peptide, the projection onto the number of native (i, i+4) hydrogen bonds is not sufficient to explain the actual complexity of the free-energy surface.

1.3.1 The protein folding network

In the last years, many complex systems, like the Internet [59], social interactions [60], metabolic pathways [61] and protein structures [62] have been

modeled as networks which proved to be a comprehensive and universal approach for the description of these systems [63]. Owing to this universality, an approach based on the theory of complex networks was developed in our group [46, 48, 64] with the purpose to overcome the limitations of projected free-energy surfaces. A network description of the energy landscape is able to retain the structural and kinetic properties of the conformations involved into the folding process. At the same time the network approach gives a qualitative two-dimensional view of the multi-dimensional free-energy surface, whose complexity is preserved. The network analysis is based on the discretization of the conformational space. In fact, a molecular dynamics trajectory is nothing but a long series of microscopic configurations visited only once. For this reason the analysis of the system needs a discretization or coarse graining of the trajectory that allows the grouping/clustering of similar snapshots/configurations. There are several procedures to efficiently achieve coarse graining and different types of analysis might require different coarse-graining approaches. For a structured peptide like α-helical peptides, secondary structural coarse-graining and root mean square deviation (RMSD) are obvious possibilities [45, 46, 58]. Both of them were used in the present work: The former combined with network and kinetic grouping analysis 1.3.2 and the latter together with cut-based free-energy profiles 1.3.3.

The conformational space network (CSN) is built as follows: conformations with similar properties are grouped together obtaining the nodes of the network. The direct transitions between them are the links [46]. Fig. 1.6 shows, as an example, the folding network of EAAAR peptide. The network reveals the presence of different regions with different α-helical content. The colorization goes from red (no helix) to green (fully formed α-helix) The

heaviest green node in the center is the most populated and its secondary structure string is -H_{15}-. A disadvantage of the network representation is that it lacks a quantitative description of the free-energy surface in the sense that no populations or barriers can be estimated from the picture. Moreover, the network contains only nodes that are populated by a significative number of snapshots (at least 50 in Fig. 1.6) to avoid overcrowding. There-

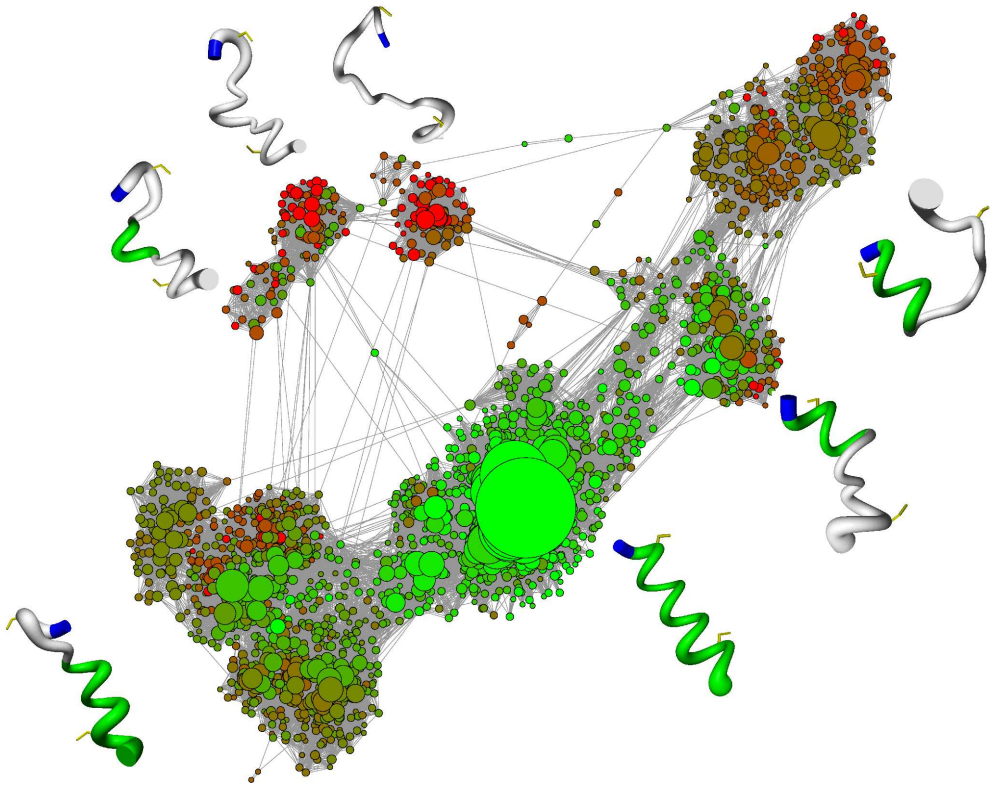
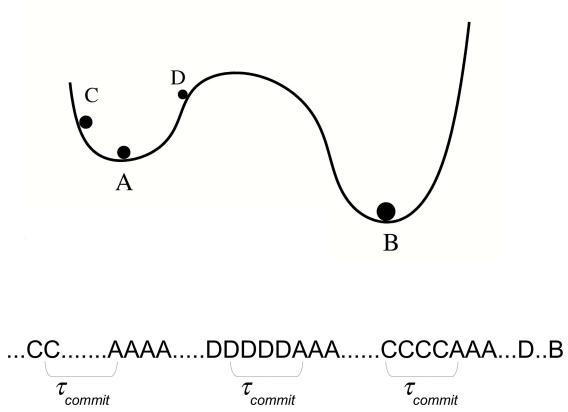


Figure 1.6: The conformational space network of EAAAR peptide at 281 K colored according to the α-helical content ranging from red to green (fully helical). Only nodes populated by at least 50 snapshots are represented to avoid overcrowding. The representative structures, made with MOLMOL [65] are shown with flexible tubes of variable diameter reflecting the conformational disorder, with α-helical segments in green, N-terminus in blue, and cysteine side chains in yellow for emphasizing the position of the cross-linker.

fore, even if the network can reveal the presence of different regions, it might miss some other aspects of the free-energy surface. With the purpose to gain quantitative information about the free-energy surface, two methods have been recently developed [66,67]. They will briefly presented in the next two sections. The common characteristic of these methods is that they are based on fast relaxation rather than the commonly used structural similarity criterion (such as the fraction of native contacts or RMSD from the folded structure) [58].

1.3.2 Kinetic Grouping Analysis

The Kinetic Grouping Analysis (KGA) method [66] is based on the observation that if two conformations interconvert rapidly they are not separated by a barrier and therefore belong to the same basin. The method requires only one parameter, the commitment time τ_{commit}, which is a typical relaxation time within the basins of the system. τ_{commit} has to be much shorter than the time required for interbasin transitions, but large enough to allow intrabasin relaxation. The probability $p_{commit}(i \rightarrow j)$ to observe a transition from node i to node j within a given τ_{commit} is an asymmetric, directed measure of the *kinetic similarity* of nodes i and j. Once the p_{commit}-matrix has been calculated for nodes containing more than a certain number of snapshots (between 50 and 500, depending on the amount of sampling), pairs of nodes (i, j) are grouped together if $p_{commit}(i \rightarrow j) \geq 0.5$. Fig. 1.7 illustrates the procedure for a simple two-state system composed of four nodes. The table on the right contains the probabilities that one node interconverts to another within the chosen τ_{commit}. The nodes that interconvert in more that 50% of the cases (i.e., nodes with interconversion probability $p_{comm} \geq 0.5$ in Fig.1.7) are grouped together (A, C, and D), whereas node B stays alone.



p_{comm}

→	A	B	C	D
A	-	0.05	0.3	0.2
B	0	-	0	0.01
C	0.9	0	-	0.5
D	0.8	0.15	0.6	-

Figure 1.7: Illustration of KGA procedure. Nodes A, C, and D interconvert rapidly within τ_{commit} and they are grouped together. For these nodes $p_{comm} \geq 0.5$. B stays alone since the barrier prevents fast relaxation between B and other nodes.

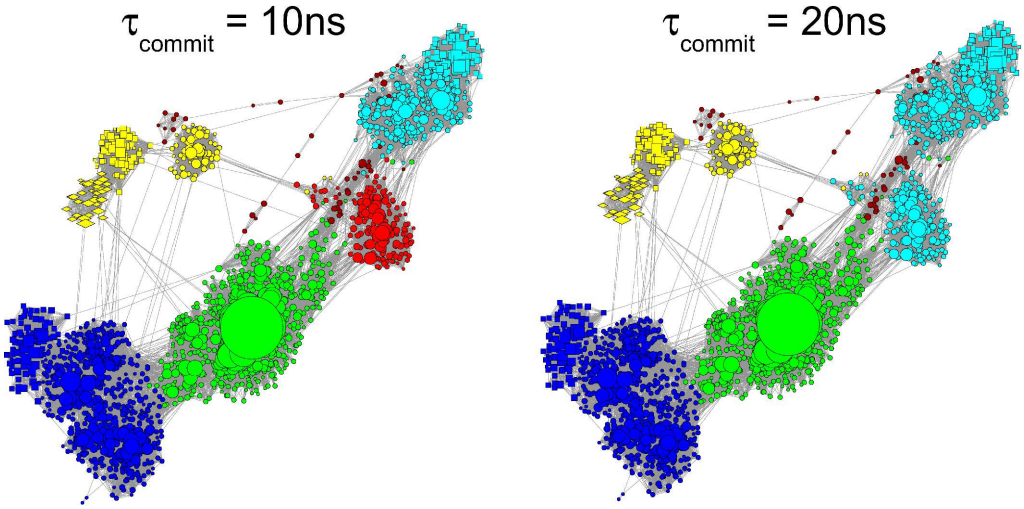


Figure 1.8: Effect of different τ_{commit}. (Left) With τ_{commit}=10 ns four main basins are obtained: green, red, blue, and cyan. (Right) When τ_{commit} is 20 ns the red basin is absorbed into the cyan one.

This way of grouping leads to a partitioning into disjoint sets, i.e., a disconnected network, whose subgraphs correspond to different basins. It is worth to mention that, different values of τ_{commit} allow one to analyze different level

of ruggedness of the free-energy surface (Fig. 1.8).

Although the kinetic grouping analysis has been introduced to analyze long equilibrium simulations, it can be also applied to non-equilibrium simulations (Fig.1.9). In fact, since kinetic grouping analysis takes into account only the rapid, local interconversion between conformations within a basin, there is no requirement for global equilibrium sampling. Therefore, also basins visited during kinetic simulations can be isolated, provided that the system equilibrates locally.

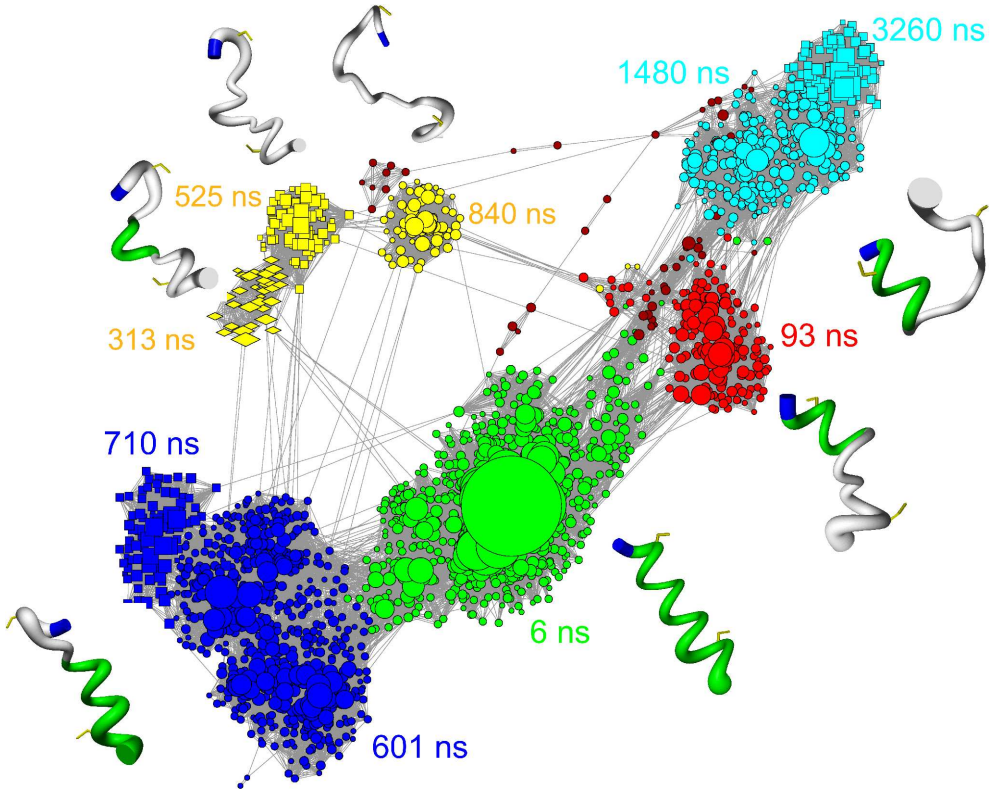


Figure 1.9: The free-energy basins of EAAAR identified by KGA with τ_{commit}=10 ns are shown with different colors. KGA shows multiple channels to the folded state (green basin). Representative structures are the same as Fig. 1.6. Brown nodes were not assigned by kinetic grouping analysis.

The network of the folding kinetics (i.e., out of equilibrium) of EAAAR presented in Fig. 1.9 shows the presence of multiple folding pathways (i.e., different basins with different color) with different time scales.

1.3.3 Cut-based Free Energy Profiles

As already discussed, the projection on one or two geometric degrees of freedom is not consistent with the complexity of the actual free-energy surface. Projected free-energy surfaces are most useful if they preserve barriers and minima in the order they are met during folding/unfolding events. A new progress coordinate, having this feature, was introduced by Krivov and Karplus [68]. The method uses the partition function of a region as the progress coordinate and determines the free-energy barriers as a function of the coordinate by a method based on the folding probability or on mean first passage time (mfpt) to a selected node. The latter was employed throughout this work. Barriers between any two nodes of interest (called representative nodes) are calculated using the minimum cut-maximal flow Ford-Fulkerson algorithm [45, 69]. For this reason the procedure is called *cut-based* free-energy profile (cFEP).

Given a network, the partition function of each node is given by

$$Z_i = \sum_i c_{ij} \tag{1.2}$$

where c_{ij} is the edge capacity from node j to node i which is proportional to the number of direct transitions from j to i. When the nodes are partitioned into two groups A and B according to the minimum cut procedure, then

$$Z_A = \sum_{i \in A} Z_i; \qquad Z_B = \sum_{i \in B} Z_i; \qquad Z_{AB} = \sum_{i \in A, j \in B} c_{ij} \tag{1.3}$$

where Z_A is the partition function of the region A, Z_B is the partition function of the region B, and Z_{AB} is the partition function of the cutting surface (i.e.,

of the barrier) that divides the cFEP into A and B (see Fig. 1.10 top panel). The free-energy of the barrier can be written as

$$\Delta G = -kTln(Z_{AB}) \tag{1.4}$$

It is possible to isolate all the basins and barriers by iterative determinations of the minimum-cuts between all pair of nodes with the Gomory-Hu algorithm [70]. In practice, to calculate the cFEP, the nodes are sorted according to their mfpt values. For any mfpt_c between 0 and mfpt_{max} a point $[Z_A/Z, -kTln(Z_{AB}/Z)]$ on the cFEP can be calculated, where A is the set of all nodes with $\text{mfpt}_i < \text{mfpt}_c$ and B the set of nodes with $\text{mfpt}_i > \text{mfpt}_c$ (a schematic illustration is presented Fig. 1.10, bottom panel).

The combination of the cFEP with a RMSD coarse-graining (i.e., a coarse-graining that groups kinetically similar snapshots) shows the presence in the helical state (i.e., trans conformation of the cross-linker) of all the photo-switchable peptides investigated of a free-energy barrier arising from two different orientations of Arg10 with respect to the cross-linker: the first has Arg10 on the right side (R-conformation, Fig. 1.11 left) whereas the second has Arg10 on the left side (L-conformation, Fig. 1.11 right). Fig.1.11 shows the cFEP of the helical state of EAAAR peptide, where the most populated node (i.e., native) has mfpt=0. All nodes lying on the left side of the barrier belong to the native basin (Z_A/Z <0.66, i.e. a population of ∼66%). Compared with secondary structure coarse-graining used for KGA analysis, RMSD coarse-graining has the main advantage that it takes into account side chain conformations. In fact, similar structures can be separated by considerable barriers, for instance if the transition between them implies the rearrangement of the side chains. This property allows one to study the R- to L-configuration transition mechanism.

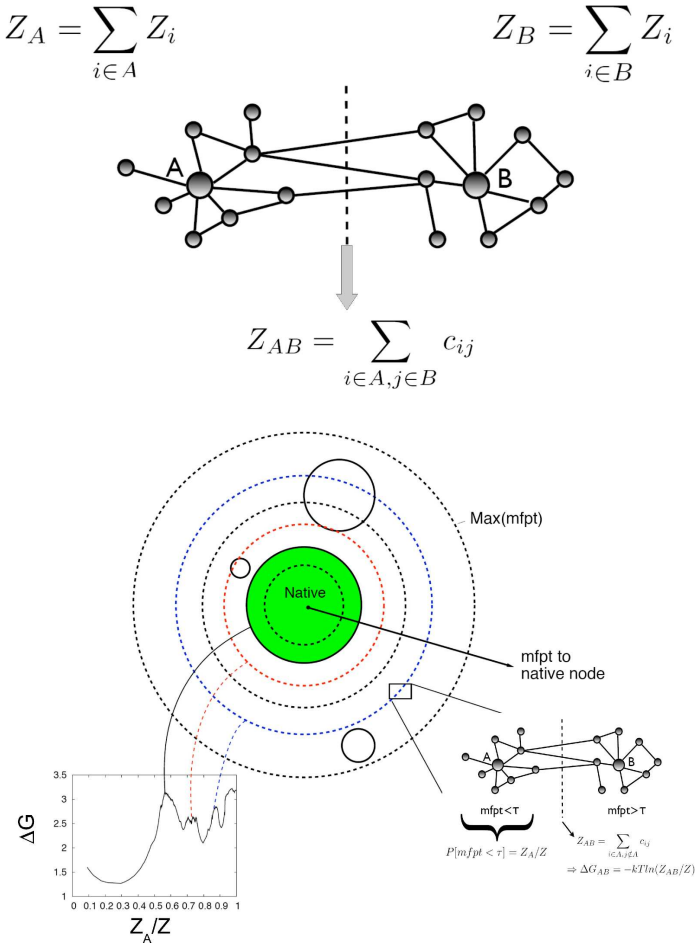


Figure 1.10: Schematic illustration of the cFEP procedure. (Top) Partitioning of the nodes in the network into to regions A and B. The cutting surface is the surface with the minimal partition function Z_{AB} that divides the cFEP into A and B. (Bottom) Schematic illustration of the one-dimensional cFEP procedure using mfpt as progress variable. Each of the four solid circles represents a free-energy basin while concentric dashed circles represent values of mfpt. For each value of mfpt_c between 0 (native node) and mfpt_{max} a point in the profile is obtained. (Bottom right) ΔG of the fraction of links crossing the cutting surface at mfpt=mfpt_c. (Bottom left) Relative partition function Z_A/Z, where the set A contains the nodes with mfpt<mfpt_c. The figure is taken from Ref. [67]

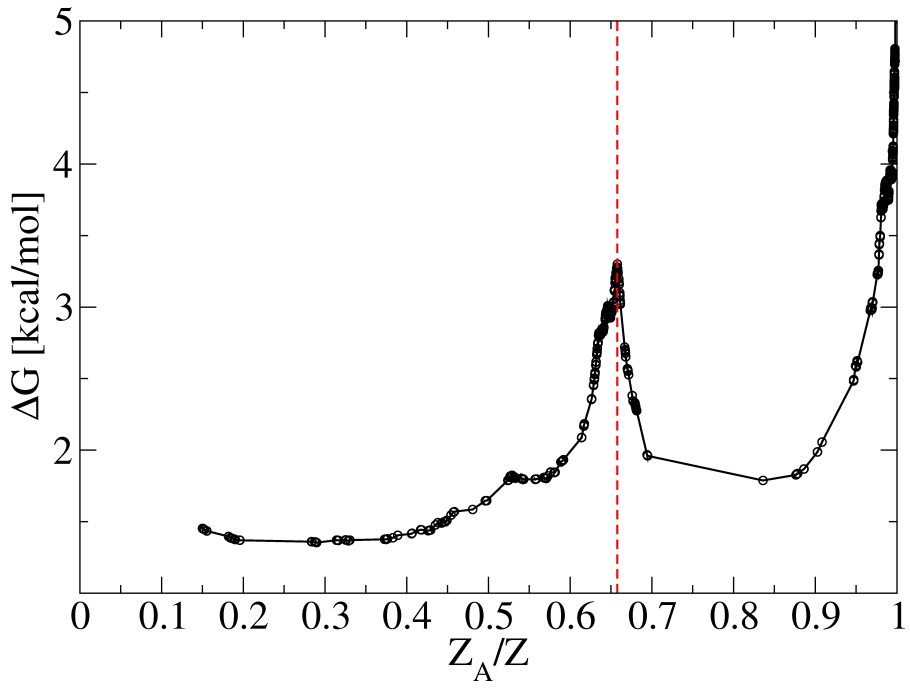


Figure 1.11: The cFEP of equilibrium trans simulations of EAAAR shows a free-energy barrier separating RMSD clusters with R-configuration of Arg10 from clusters with L-configuration. The height of the barrier is about 2 kcal/mol. The representative structure on the left side of the barriers has R-conformation of Arg10, whereas the representative structure on the right side has L-configuration.

1.4 Thesis Scheme

This thesis is organized with the following structure consisting in the corresponding publications:

- **Chapter 3.** ***α-Helix Folding in the Presence of Structural Constraints.*** The study of the site-specific folding kinetics of AAAAR peptide from both experimental and computational sides, shows that the folding times differ form site to site at low temperature (281 K) and moreover different sites fold in a noncooperative manner. The combination of network analysis and KGA to analyze the MD simulations shows

that the unfolded state is kinetically partitioned into parallel folding routes having different folding times. Moreover, the barriers within the unfolded state are higher that barriers between unfolded basins and the native one, providing evidence that the native state acts as a hub.

- **Chapter 4.** ***Bulky side chains and non-native salt bridges slow down the folding of a cross-linked helical peptide: A combined molecular dynamics and time-resolved infrared spectroscopy study.*** This work shows that the network analysis of the MD simulations in combination with KGA provide a description of the free-energy surface of α-helix formation that not only captures the complex folding behavior and sequence dependence but can also be used to make suggestions for guiding experiments.

- **Chapter 5.** ***Cross-linked helical peptides have complex free-energy surface.*** In this paper we investigate the role of the entanglement of bulky side chains with the cross-linker using the cFEP approach, combined with side chains RMSD coarse-graining. The cFEP of the helical state shows a barrier arising from two different orientations of Arg10 with respect to the cross-linker. Additionally, we compared the behavior of the folding kinetics of AAAAR, EAAAR and EMAAAR in three different cases: in the presence of the cross-linker; when the cross-linker is substituted with a distance constraint mimicking the effect of a "non-bulky" cross-linker; in the absence of constraints (free-system).

Bibliography

[1] C. B. Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181:223–220, 1973.

[2] C. Levinthal. Are there pathways for protein folding? *J. Chim. Phys.*, 65:44–65, 1968.

[3] K. Dill and H.S. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.*, 4:10–18, 1997.

[4] R. L. Baldwin. Matching speed and stability. *Nature*, 369:183–184, 1994.

[5] A. Šali, E. Shakhnovich, and M. Karplus. How does a protein fold. *Nature*, 369:248–251, 1994.

[6] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21:167–195, 1995.

[7] M. Karplus. The Levinthal paradox: yesterday and today. *Folding Des.*, 2:S69–S65, 1997.

[8] P.E. Leopold, M. Montal, and J.N. Onuchic. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, 89:8721–8725, 1992.

[9] P.G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–1620, 1996.

[10] P.G. Wolynes. Symmetry and the energy landscapes of biomolecules. *Proc. Natl. Acad. Sci. USA*, 93:14249–14255, 1996.

[11] K.A. Dill. *The stabilities of globular proteins.* D.L. Oxender and C.F. Fox, 1987.

[12] O.B. Ptitsyn and A.A. Rashin. A model of myoglobin self-organization. *Biophys. Chem.*, 3:1–20, 1975.

[13] P. S. Kim and R. L. Baldwin. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.*, 59:631–660, 1990.

[14] B. Nölting and K. Andert. Mechanisms of protein folding. *Proteins: Structure, Function, and Genetics*, 41:288–298, 2000.

[15] S. Rackovsky and H. A. Scheraga. Hydrophobicity, hydrophilicity, and the radial and orientational distributions of residues in native proteins. *Proc. Natl. Acad. Sci. USA*, 74:5248–5251, 1977.

[16] K.A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501–1509, 1985.

[17] D.B. Wetlaufer. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*, 70:697–701, 1973.

[18] A. R. Fersht. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA*, 92:10869–10873, 1995.

[19] P.A. Evans, K.D. Topping, D.N. Woolfson, and C.M. Dobson. Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins: Structure, Function, and Bioinformatics*, 9:248–266, 1991.

[20] M. Buck, S.E. Radford, and C.M. Dobson. Amide hydrogen exchange in a highly denatured state: hen egg-white lysozyme in urea. *J. Mol. Biol.*, 237:247–254, 1994.

[21] H. Schwalbe, K.M. Fiebig, M. Buck, J.A. Jones, S.B. Grimshaw, A. Spencer, S.J. Glaser, and L.J. Smith nad C.M. Dobson. Structural and dynamical properties of a denatured protein. heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8M urea. *Biochemistry*, 36:8977–8991, 1997.

[22] D. Neri, M. Billeter, G. Wider, and K. Wüthrich. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, 257:1559–1563, 1992.

[23] T.M. Logan, Y. Thèriault, and S.W. Fesik. Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.*, 236:637–348, 1994.

[24] K.J. Lumb and P.S. Kim. Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, 236:412–420, 1994.

[25] M.K. Frank, G.M. Clore, and A.M. Gronenborn. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of Streptococcal protein G. *Protein Science*, 4:2605–2615, 1995.

[26] V.L. Arcus, S. Vuilleumier, S.M.V. Freund, M. Bycroft, and A.R. Fersht. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.*, 254:305–321, 1995.

[27] O. Zhang and J.D. Forman-Kay. structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry*, 34:6784–6794, 1995.

[28] O. Zhang and J.D. Forman-Kay. NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry*, 36:3959–3970, 1997.

[29] J.R. Gillespie and D. Shortle. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.*, 268:158–169, 1997.

[30] J.R. Gillespie and D. Shortle. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. distance restraints from paramagnetic relaxation and calculation of an ensemble of structures'. *J. Mol. Biol.*, 36:170–184, 1997.

[31] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, and H.E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.

[32] A. Engel, H.E. Gaub, and D.J. Müller. Atomic force microscopy: a forceful way with single molecules. *Curr. Biol.*, 9:R133–R136, 1999.

[33] B. Schuler. Single-molecule fluorescence spectroscopy of protein folding. *Chemphyschem*, 6:1206–1220, 2005.

[34] S. Williams, T.P. Causgrove, R. Gilmanshin, K.S. Fang, R.H. Callender, W.H. Woodruff, and R.B. Dyer. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, 35:691–697, 1996.

[35] C.Y. Huang, J.W. Klemke, Z. Getahun, W.F. DeGrado, and F. Gai. Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc.*, 123:9235–9238, 2001.

[36] C.Y. Huang, Z. Getahun, Y. Zhu, J.W. Klemke, W.F. DeGrado, and F. Gai. Helix formation via conformation diffusion search,. *Proc. Natl. Acad. Sci. USA*, 99:2788, 2002.

[37] J.H. Werner, R.B. Dyer, R.M. Fesinmeyer, and N.H. Andersen. Dynamics of the primary processes of protein folding: helix nucleation. *J. Phys. Chem. B*, 106:487–494, 2002.

[38] S.A. Petty and M. Volk. Fast folding dynamics of an α-helical peptide with bulky side chains. *Phys. Chem. Chem. Phys.*, 6:1022–1030, 2004.

[39] P.A. Thompson, W.A. Eaton, and J. Hofrichter. Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a "kinetic zipper" model. *Biochemistry*, 36:9200–9210, 1997.

[40] P.A. Thompson, V. Muñoz, G.S. Jas, E.R. Henry, W A. Eaton, and J. Hofrichter. The helix-coil kinetics of a heteropeptide. *J. Phys. Chem. B*, 104:3278–389, 2000.

[41] I.K. Lednev, A.S. Karnoup, M.C. Sparrow, and S.A. Asher. α-helix peptide folding and unfolding activation barriers: A nanosecond UV resonance raman study. *J. Am. Chem. Soc.*, 121:8074–8086, 1999.

[42] E. R. Henry and W. A. Eaton. Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem. Phys.*, 307:163–185, 2004.

[43] A. N. Naganathan, U. Doshi, A. Fung, M. Sadqi, and V. Muñoz. Dynamics, Energetics, and Structure in Protein Folding. *Biochemistry*, 45:8466–8475, 2006.

[44] G. Hummer, A. E. Garcia, and S. Garde. Helix nucleation kinetics from molecular simulations in explicit solvent. *Proteins*, 42:77–84, 2001.

[45] S.V. Krivov and M. Karplus. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. USA*, 101:14766–14770, 2004.

[46] F. Rao and A. Caflisch. The protein folding network. *J. Mol. Biol.*, 342:299–306, 2004.

[47] Y. Mu, P. Nguyen, and G. Stock. Energy landscape of a small peptide revealed by dihedral angle principle component analysis. *Proteins*, 58:343–357, 2005.

[48] A. Caflisch. Network and graph analyses of folding free energy surfaces. *Curr. Opin. Struc. Biol.*, 16:71–78, 2006.

[49] J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, and P. Hamm. α-Helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 102:2379–2384, 2005.

[50] P. Ferrara, J. Apostolakis, and A. Caflisch. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 46:24–33, 2002.

[51] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[52] J.R. Kumita, O.S. Smart, and G.A. Woolley. Photo-control of helix content in a short peptide. *Proc. Natl. Acad. Sci. USA*, 97:3803–3808, 2000.

[53] A.R. Dinner, A.Šali, L.J. Smith, C.M. Dobson, and M.Karplus. Understanding protein folding via free-energy surfaces form theory and experiments. *Trends in Biochemical Science*, 25:331–339, 2000.

[54] L.A. Mirny and E.I. Shakhnovich. Protein folding theory: from lattice to all-atom models. *Ann. Rev. Biophys. Biomolec. Struct.*, 30:361–396, 2001.

[55] V. Daggett and A.R. Fersht. The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell. Biol.*, 4:497–502, 2003.

[56] P.G. Wolynes. Enery landscapes and solved protein-folding problems. *Phil. Trans. R. Soc. A*, 363:453–467, 2005.

[57] R. Du, V.S. Pande, A.Y. Grosberg, T. Tanaka, and E.S. Shakhnovich. On the transition coordinate for protein folding. *J. Chem. Phys.*, 108:334–350, 1998.

[58] I.A. Hubner, E.J. Deeds, and E.I. Shakhnovich. Understanding ensemble protein folding at atomic detail. *Proc. Natl. Acad. Sci. USA*, 103:17747–17752, 2006.

[59] R. Albert, H. Jeong, and A.L. Barabasi. Internet-diameter of the world-wide web. *Nature*, 401:130–131, 1999.

[60] M.E.J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98:404–409, 2001.

[61] H. Jeong, B. Tombor, R. Albert, Z.N. Oltval, and A.L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[62] L.H. Greene and V.A. Higman. Uncovering network systems within protein structures. *J. Mol. Biol.*, 334:781–791, 2003.

[63] M.E.J. Newman. The structure and function of complex networks. *Siam Review*, 45:167–256, 2003.

[64] D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao. Complex network analysis of free-energy landscape. *Proc. Natl. Acad. Sci. USA*, 104:1817–1822, 2007.

[65] R. Koradi, M.Billeter, and K. Wüthrich. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics Modell.*, 14:51–55, 1996.

[66] S. Muff and A. Caflisch. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein. *Proteins: Structure, Function, and Bioinformatics*, 70:1185–1195, 2008.

[67] S.V. Krivov, S. Muff, A. Caflisch, and M. Karplus. One-dimensional barrier-preserving free-energy projections of a β-sheet miniprotein: new

insights into the folding process. *J. Phys. Chem. B*, 112:8701–8714, 2008.

[68] S.V. Krivov and M. Karplus. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B*, 110:12689–12698, 2006.

[69] L.R. Ford and D.R. Fulkerson. Maximal flow through a network. *Canadian J. of Math.*, 8:399–404, 1956.

[70] R.E. Gomory and T.C. Hu. Multi-terminal network flows. *J. Soc. Indust. Appl. Math.*, 9:551–570, 1961.

Chapter 2

α-helix folding in the presence of structural constraints

[*PNAS*, **2008**, *105*, 9588-9593]

α-Helix folding in the presence of structural constraints

Janne A. Ihalainen*†, Beatrice Paoli†‡, Stefanie Muff‡, Ellen H. G. Backus*, Jens Bredenbeck*, G. Andrew Woolley§, Amedeo Caflisch‡¶, and Peter Hamm*¶

*Physikalisch-Chemisches Institut and ‡Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; and §Department of Chemistry, University of Toronto, 80 Saint George Street, Toronto, ON, Canada M5S 3H6

Edited by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved April 23, 2008 (received for review December 22, 2007)

We have investigated the site-specific folding kinetics of a photoswitchable cross-linked α-helical peptide by using single $^{13}C = {}^{18}O$ isotope labeling together with time-resolved IR spectroscopy. We observe that the folding times differ from site to site by a factor of eight at low temperatures (6°C), whereas at high temperatures (45°C), the spread is considerably smaller. The trivial sum of the site signals coincides with the overall folding signal of the unlabeled peptide, and different sites fold in a noncooperative manner. Moreover, one of the sites exhibits a decrease of hydrogen bonding upon folding, implying that the unfolded state at low temperature is not unstructured. Molecular dynamics simulations at low temperature reveal a stretched-exponential behavior which originates from parallel folding routes that start from a kinetically partitioned unfolded ensemble. Different metastable structures (i.e., traps) in the unfolded ensemble have a different ratio of loop and helical content. Control simulations of the peptide at high temperature, as well as without the cross-linker at low temperature, show faster and simpler (i.e., single-exponential) folding kinetics. The experimental and simulation results together provide strong evidence that the rate-limiting step in formation of a structurally constrained α-helix is the escape from heterogeneous traps rather than the nucleation rate. This conclusion has important implications for an α-helical segment within a protein, rather than an isolated α-helix, because the cross-linker is a structural constraint similar to those present during the folding of a globular protein.

cooperativity | infrared spectroscopy | molecular dynamics simulation | peptide folding

In many biomolecular systems, large changes can take place in response to a relatively small perturbation in the environment such as a variation in temperature, denaturant concentration, or the partial pressure of certain gases. Such an "all or nothing" phenomenon is termed a cooperative process. The classical example of cooperativity is the binding affinity of oxygen molecules to the four hemes of hemoglobin (1), which is a factor 100 to 1,000 times larger for the fourth oxygen molecule compared with the first. This leads to a sigmoidal dependence of oxygen binding on oxygen partial pressure with a sharp transition in a relatively small range of the latter. The folding of α-helices, which constitute one of the predominant secondary structures in many proteins, is often described in a similar manner (2): Once an entropically expensive nucleation process has occurred, i.e., a first helical turn with a hydrogen bond is formed, the zipping of additional hydrogen bonds is more likely because it is enthalpically favorable. A thermodynamic (statistical) treatment of the process leads to so-called nucleation-propagation-models (or zipper models), initially introduced by Zimm and Bragg (3) and Lifson and Roig (4). A large number of thermodynamic studies on α-helical peptides has been treated extremely successfully in terms of these models (5–8).

Cooperativity implies that the free energies of the two states of a system are balanced, however, in a way that enthalpic (ΔH) and entropic ($-T\ \Delta S$) contributions are large and compete against each other, leading to a characteristic sigmoidal transition as a function of the external control parameter. Because, in general, enthalpy and entropy vary in a nonsynchronous way as a function of some order parameter, the resulting free-energy surface $\Delta G \equiv \Delta H - T\Delta S$ will be uneven and in most cases will have a pronounced barrier. This is why this definition of cooperativity, which is based on thermodynamic arguments, often also has consequences for the kinetics of the transition between the two states. For example, in the classic case of oxygen binding to hemoglobin, a two-state allosteric model (i.e., the MWC model) can also explain the binding rate that increases with the number of already bound oxygen molecules (1). Cooperativity in protein folding reflects a two-state conformational distribution; its investigation requires a rigorous analysis of the folding transition (9). In the case of the helix–coil transition, high cooperativity would imply that once the rate-limiting nucleation step has occurred somewhere in the sequence, all subsequent helical turns would form at essentially the same time. Then, one common rate would be expected for all sites, corresponding to the nucleation rate (the propagation rate would not be detectable because it is very fast). Indeed, temperature-jump experiments on helix folding have successfully been described by "kinetic-zipper" models (10, 11), in which thermodynamic states of a nucleation-propagation model are linked by rate constants. However, because the cooperativity of isolated α-helices is weak, in particular when they are short, they do not fold in a two-state fashion, but rather with biexponential kinetics resulting from the coupling between nucleation and the only slightly faster diffusive elongation (10, 12).

It has recently been argued that even proteins that appear to be two-state folders can in fact be much more complex when reporting the folding free-energy surface on the level of individual protons by using NMR chemical-shift spectroscopy (13). IR spectroscopy together with site-selective isotope labeling offers the time resolution to perform site-selective folding studies also in a kinetic sense, even for the much faster folding of secondary structure motifs. The amide I′ (C = O stretch) vibrational mode is very sensitive to hydrogen bonding and dipole–dipole coupling among different peptide units (14), and isotope labeling of the carbonyl groups allows one to spectrally single out individual amino acids (12, 15). Gai and coworkers (12) demonstrated by $^{13}C = {}^{16}O$ labeling of groups of four

Author contributions: G.A.W., A.C., and P.H. designed research; J.A.I., E.H.G.B., and J.B. performed experiments; B.P. and S.M. performed simulations; G.A.W. contributed new reagents/analytic tools; J.A.I., B.P., S.M., E.H.G.B., J.B., A.C., and P.H. analyzed data; and J.A.I., A.C., and P.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†J.A.I. and B.P. contributed equally to this work.

¶To whom correspondence may be addressed. E-mail: p.hamm@pci.uzh.ch or caflisch@bioc.uzh.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0712099105/DCSupplemental.

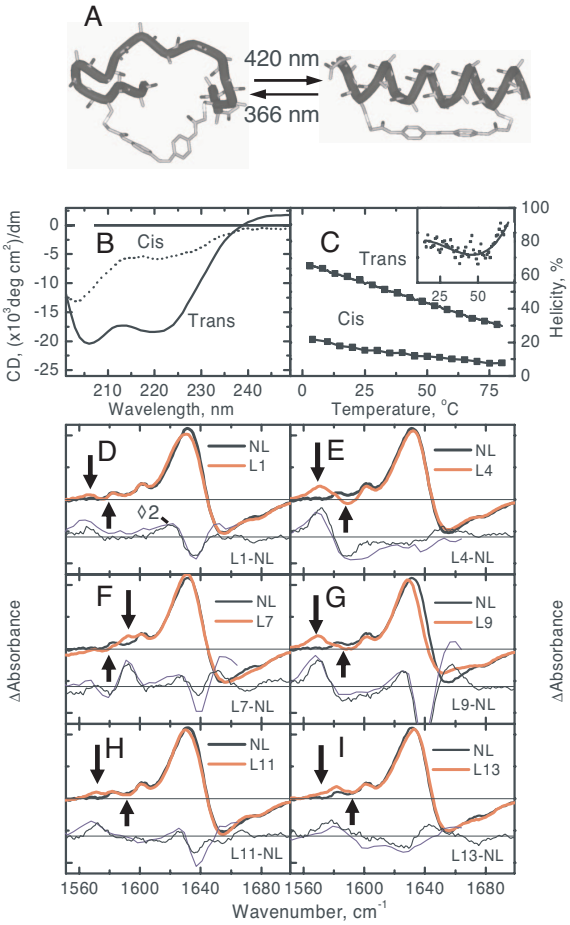


Fig. 1. Photoswitchable peptide and its steady-state spectra at room temperature. (*A*) Schematic drawing of the photoswitchable peptide in its *cis* (*Left*) and *trans* (*Right*) conformations. (*B*) CD-spectra of the peptide in its complete *trans* and *cis* states at room temperature. (*C*) The helicities in the *cis* and *trans* conformations at various temperatures. The *trans* spectrum was obtained in darkness, and the *cis*-spectrum was obtained under 365-nm illumination together with an estimate that 75% of the molecules are in the *cis* conformation (based on the UV-vis absorbance difference). (*Inset*) The first derivative of the *trans* CD data reveals the inflection point. (*D–I*) FTIR-difference spectra between the *trans* (under 436-nm illumination) and *cis* state (under 365-nm illumination) of each peptide at room temperature. Shown are the difference between not-labeled (NL) and labeled (L*X*, labeled at residue *X*) samples taken from the FTIR-spectra (thin black line) and from the late delay-time pump-probe spectra 30 μs after photoswitching (thin blue line). The arrows point to the frequency position from which the site-specific kinetics has been extracted.

subsequent alanines that the relaxation rates vary by ≈10% throughout the sequence. This result points to a more complex behavior of α-helix folding. Nevertheless, such a small degree of heterogeneity could be still modeled in the framework of nucleation-propagation models, taking into account position-dependent parameters for the different amino acids in the heteropolymer (16). However, as we will show, labeling of groups of several subsequent amino acids might still be too coarse-grained. If the kinetics in various sections of the peptide differ, one might expect variations even on the single amino acid level. Indeed, by using $^{13}C = {}^{18}O$ double labeling of a single amino acid in a small helix bundle protein, Dyer and coworkers (17) obtained considerably different melting curves and considerably different complex folding kinetics for that particular site, compared with the averaged signal.

As an alternative approach to study α-helix folding, we recently started to employ an azobenzene moiety as a photoswitchable structural constraint (see Fig. 1*A*) (18–20). Two cysteines are cross-linked in such a way that the azo-moiety in the *trans* (*cis*) conformation stabilizes (destabilizes) the helix, as deduced from CD spectroscopy (Fig. 1 *B* and *C*). This allows one to monitor both the folding and the unfolding direction of one and the same molecule at identical temperatures. We observed stretched exponential kinetics in both directions (20), in disagreement with the prediction from nucleation-propagation models that would reveal compressed exponential kinetics in the folding direction (21) (in fact, compressed exponential response has already been obtained in figure 10 of ref. 10, however, remained undiscussed). We therefore speculated in our previous work that the rate-limiting step in folding of our model systems is the escape from misfolded traps, which are consistently obtained in full-atom molecular dynamic (MD) simulations of small peptides (22–29).

Here, we present a comprehensive set of kinetic data of the folding of a photoswitchable α-helix (Ac-AACAK^{5}AAAAK^{10}AAACK^{15}A-NH$_2$) on the single amino acid level by employing $^{13}C = {}^{18}O$ labeling of an alanine-rich peptide, revealing strong variations in the folding kinetics and their temperature dependence along the sequence. The interpretation of the experimental data are corroborated by multiple implicit solvent (30) MD simulations (31) of folding of a similar cross-linked peptide (i.e., Ac-AACAR^{5}AAAAR^{10}AAACR^{15}A-NH$_2$) at low and high temperatures, and their network analysis (22, 29). Nonequilibrium MD simulations of an eight-residue cyclized peptide immersed in dimethyl sulfoxide have been published recently (32, 33) but those studies focused on the contributions to the frequency shift and did not mention pathways and kinetics of (helical) folding. The atomistic detail of the MD simulations and the large sampling of folding events (100 MD runs for a total of 0.4 ms) allow us to directly extract the conformational distribution from the trajectories. The comparison of experimental data and simulation results shows qualitatively similar folding kinetics and temperature dependence. Our analysis provides strong evidence that (*i*) the azo-cross-linker stabilizes kinetic traps originating from nonnative contacts that are nonexistant in isolated helices and that (*ii*) the azo-cross-linker finally destroys the already weak cooperativity of isolated α-helices. The analysis of the photoswitching simulations provides insights, at the atomic level of detail, on the folding mechanism and thereby explains the kinetic traces measured experimentally.

Results and Discussion

Noncooperative Helix Folding. Fig. 1*C* shows melting curves deduced from the fractional helicity at 222 nm [assuming $[\Theta]_{222} = -32{,}000$ deg cm^2/dm for the ellipticity of a fully folded helix (34)]. Within the limited temperature range, the sigmoidal dependence of the folding transition is hardly visible (see Fig. 1*C* *Inset*). Nevertheless, the helicity in the *trans* conformation varies slightly more than the relative temperature change ($\Delta T/T$ = 24%), hence we conclude that some small degree of cooperativity still remains, in the sense that folding enthalpy and folding entropy compete against each other. A comparison with melting curves from similar but nonlinked peptides (34, 35), which are steeper, indicates that the linker reduces the conformational flexibility of the unfolded ensemble and, hence, diminishes the entropic penalty of folding. It should be noted that the folding rates we observe for the azo-linked peptides (19, 20) are similar to those found in temperature-jump experiments of comparable unlinked peptides (10, 12, 36, 37) (a one-to-one comparison is

not possible because the sequences used in temperature-jump experiments are typically a bit longer) and that the amino acid sequence determines the folding rate to a significant extent. That is, at room temperature, the folding rate is 1,200 ns for Ac-EACAR^{5}EAAAR^{10}EAACR^{15}Q-NH$_2$ (19), 700 ns for Ac-AACAR^{5}AAAAR^{10}AAACR^{15}A-NH$_2$ (20), 600 ns for Ac-AACAK^{5}AAAAK^{10}AAACK^{15}A-NH$_2$ (this study), and 2,200 ns for Ac-EMCAR^{5}EMAAR^{10}EMACR^{15}Q-NH2 (data not shown). This variability indicates that interactions among the amino acid side chains and with the cross-linker may act as traps that are, indeed, rate determining.

Stationary FTIR-difference spectra between the two conformations of the photoswitchable helix are shown in Fig. 1 *D–I*. Upon folding, the unlabeled amide I′ band red-shifts with a positive signal ≈1633 cm^{-1} and a negative signal with additional substructure ≈1655 cm^{-1} and 1680 cm^{-1}, reporting on an overall strengthening of hydrogen bonding (14). Additional small bands are observed at lower frequencies that originate from both the ring modes of the azo-moiety (1,602 cm^{-1} and 1,580 cm^{-1}) (20) and from the isotope labels. By subtracting the FTIR-difference spectrum of the nonlabeled compound (termed NL throughout the text) from that of isotope-labeled compounds (termed L*X*, where *X* is the labeling position counted from the N terminus), the contributions of the latter can be isolated. As expected, they are downshifted by ≈65 cm^{-1} from the main band, exhibiting a dispersive shape with a sharp and distinct positive contribution in either case, whereas the negative contribution is, in general, broader and less clearly identified (see arrows in Fig. 1 *D–I*). The one notable exception is L7, which shifts to higher frequencies upon folding of the α-helix (Fig. 1*F*), opposite to all other residues. Because the frequency of the ^{13}C = ^{18}O vibration is related to the strength of hydrogen bonding, the only conclusion can be that hydrogen bonding of this particular site is stronger in the "unfolded" ensemble.

In the time-resolved experiments, helix folding is initiated by a subpicosecond laser pulse isomerizing the azo-photoswitch, and the formation of individual hydrogen bonds is detected in the amide I′ region as a function of time (19, 20). To single out the contribution from the isotope label, difference spectra between nonlabeled and labeled samples had to be taken. Furthermore, to suppress temperature-induced baseline effects, two kinetics were collected for each sample (Fig. 2 *A* and *B*; NL, open circles; L1, filled circles) at the frequency positions with the biggest deviations between nonlabeled and labeled sample (left and right from the ^{13}C = ^{18}O vibration, as indicated by arrows in Fig. 1). The result of subtracting the normalized signals from NL and L1 is shown in Fig. 2*C* (normalization was performed according to concentration obtained from the UV-vis spectrum and excitation power), and the difference between the traces at the two frequencies is shown in Fig. 2*D*. This "difference of difference signal" constitutes the site-selective folding trace and can be fitted (within signal-to-noise) with a single-exponential function shown as a solid line in Fig. 2*D* for L1. We note also that the difference–difference signals of the late time transient experiment agree with that from the FTIR-experiment (Fig. 1), validating the correctness of the background subtraction procedure. The site-selective folding traces labeled at all other positions are shown in Fig. 2 *E–J*. The immediate observation is that the rates scatter quite substantially without clear correlation between neighboring sites (Fig. 3*A*), thus, each peptide unit behaves independently, or noncooperatively. As a general trend, the signal is larger in the middle of the peptide, and sites 4, 7, and 9 reveal the most dominant signals (however, note again that L7 exhibits kinetics in the opposite direction; Fig. 2*F*). In line with other studies (e.g., ref. 38), a so-called "end-fraying effect" is observed as evidenced by smaller signals at the C and N termini. As a matter of fact, one would not expect any response from L13 and L16 at all, because they cannot form α-helical hydrogen bonds even in the folded conformation, and indeed, the signals

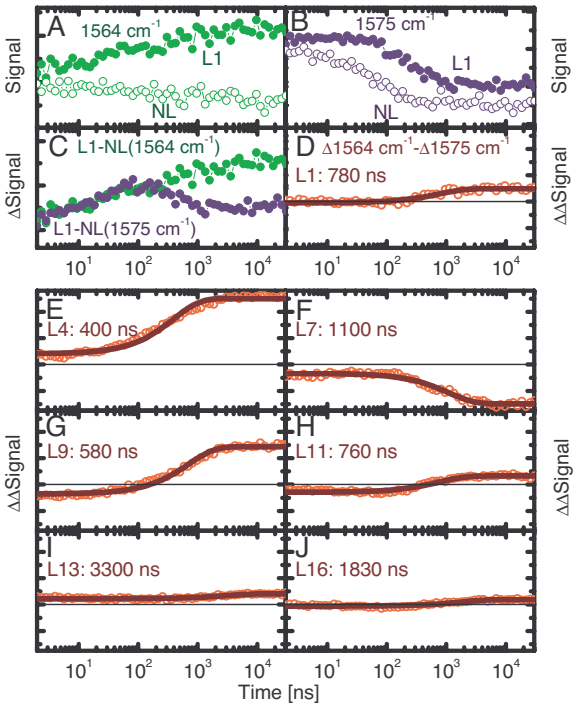


Fig. 2. Site-specific folding signals at 19°C. (*A* and *B*) Kinetic traces of the L1 and NL at 1,564 cm^{-1} (*A*) and at 1575 cm^{-1} (*B*). (*C*) The difference of the L1 and NL signals at these wavelengths. (*D*) The difference between the resulting signals together with its single exponential fit (solid line). (*E–J*) Site-selective folding signals of the other sites. The solid lines are single-exponential fits of the curves with the resulting time constant indicated.

are extremely small. Interestingly, the rates are correlated with the amplitudes of the corresponding signals (Fig. 3*A*); thus, the larger the driving force, the faster hydrogen bond formation.

Temperature Dependence. It is useful to plot the rates of folding of individual sites (i.e., the rates of helical hydrogen bond formation of individual backbone carbonyl groups) against reciprocal temperature, as measured by the isotopically labeled peptides (Fig. 3*C*). The individual sites exhibit considerably different temperature dependencies. The spread of rates is clearly bigger at low temperatures (a factor of approximately eight at 6°C), whereas they approach each other at higher temperatures (a factor of two at 45°C). Interestingly, summing up all site-signals with their relative intensities, the kinetics of the unlabeled band can be reproduced remarkably well (Fig. 3*E*). This result shows that our set of single-site labeled peptides represents the overall folding kinetics very well and, more importantly, that the averaged kinetics is just a trivial sum of the individual contributions. The average folding time of the NL peptide increases from approximately $\tau = 240$ ns to $\tau = 1{,}290$ ns when lowering the temperature from 45°C to 6°C and becomes more stretched with a stretching factor that decreases from $\beta = 1.00$ to $\beta = 0.71$ (fitting it with a function $\propto \exp[-(t/\tau)^{\beta}]$) (19, 20). The larger spread of rates reveals stretched exponential kinetics at low temperatures, whereas the more uniform values at higher temperatures result in a close to single-exponential response.

Qualitatively speaking, the MD simulations reveal very similar results. Just as in the experiment, the folding rates and amplitudes vary strongly from site to site (Fig. 3*B*), with a fraying effect toward the ends and a dip in the folding amplitudes in the middle

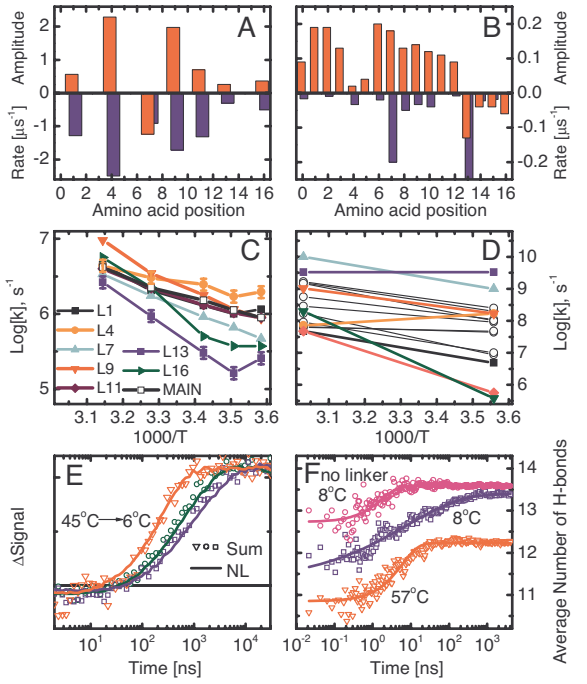


Fig. 3. Summary of the experimental (*A*, *C*, and *E*) and MD (*B*, *D*, and *F*) results. (*A* and *B*) Amplitudes (upward, red) and corresponding rates (downward, blue) at 19°C (*A*) and 8°C (*B*). Note that site 7 in the experiment, and sites 13–16 in the MD simulations, have inverted amplitudes. (*C* and *D*) Site-selective folding rates as a function of inverse temperature. In *D*, black circles are used for residues not measured experimentally. (*E*) Sum of all site signals (symbols) and amide I′ signal of the nonlabeled peptide (solid lines) at 6°C, 19°C, and 45°C. (*F*) Average number of hydrogen bonds along the MD simulations: 100 MD runs at 8°C with stretched exponential fit $h(t) = 13.6 - 1.6\exp(-t/29\ \text{ns})^{0.39}$ (blue), 50 MD runs at 57°C with single-exponential fit $h(t) = 12.2 - 1.4\exp(-t/6\ \text{ns})$ (red), and 50 MD runs of peptide without cross-linker at 8°C with single-exponential fit $h(t) = 13.6 - 0.9\exp(-t/2\ \text{ns})$ (magenta). The standard deviations are almost all <0.3 units. The experimental data in *E* are normalized, and the background has been removed, whereas the corresponding MD data in *F* are on their original scale.

of the helix (albeit not exactly at the same position). Also the temperature dependence observed in the MD simulations is in qualitative agreement with the experimental data (Fig. 3*D*), despite the larger spread of individual-residue rates, which is likely to originate mainly from the low friction coefficient [see supporting information (SI) *Text*]. Quantitative agreement is not expected because of the approximations inherent to the force field and implicit solvation model as well as the slightly different amino acid sequence. As an example, the fastest rate observed in MD (L13, blue square symbol in Fig. 3*D*) is the slowest in the experiment (Fig. 3*C*), which is in part a consequence of the very small amplitude. Nevertheless, essentially the same overall folding kinetics and temperature dependence emerge from the analysis of the MD simulations in terms of total number of hydrogen bonds involving carbonyl groups (Fig. 3*F*), which is the MD observable closest to the experimental signal. The folding signal from the MD runs at low temperature shows complex kinetics, clearly deviating from single-exponential behavior.

There are indications for deviations from single-exponential response even on the single-site level in both the experimental (Fig. 2) and the MD data (Fig. S3 and Tables S2 and S3), however, because of limited signal-to-noise, we do not discuss them in detail. Nevertheless, the largest contribution to the nonexponentiality of the overall signal originates from the spread of rates rather than from the nonexponentiality of the individual signals. Moreover, just as in the experiment, faster folding and single-exponential behavior are observed in the MD runs at high temperature (Fig. 3*F*, red). In particular, the ratio of folding times at low vs. high temperature is close: 29:6 in the simulations of the cross-linked peptide and 1,290:240 as measured experimentally. Finally, we simulated folding of the peptide without the cross-linker (Fig. 3*F*, magenta, an "experiment" that can only be performed on the computer) and obtain single-exponential response even at low temperature.

Origin of Complex Kinetics Explained by Network Analysis of MD Trajectories. In the example of oxygen binding to hemoglobin, cooperativity implies that the first step is the rate-limiting step (1). If the same were true for the folding of our photoswitchable α-helix, one common rate would be expected for all sites corresponding to the rate-limiting nucleation step. Such a common rate is not observed; instead, each amino acid site responds individually, reporting on various escape rates from different partially folded or trap states. The averaged signal obtained from the main band is then a nonspecific sum of these rates, and hence, the folding of the cross-linked α-helix is definitely not cooperative in the kinetic sense, although, from thermodynamic considerations (Fig. 1*C*), one might still deduce some degree of cooperativity. The complexity of the folding process is masked in an averaged signal, as suggested by previous atomistic simulation studies (22, 23), which have demonstrated that projecting the free energy on a single progress variable based on geometry, e.g., number of native contacts or rmsd from native, is not consistent with the complexity of the actual free-energy surface (25). Note that these projections (histogram-based free-energy profiles in Fig. S4) do not reveal any rate-limiting barrier for the photoswitchable α-helix.

The agreement of the MD results with the experiment justifies drawing detailed conclusions from the former. To that end, the folded (i.e., fully α-helical) state, and the most populated metastable states, can be isolated by grouping conformations according to fast relaxation along the MD trajectories, a procedure called kinetic grouping analysis (29). The advantage of this procedure with respect to a simple projection onto one or two progress variables is that structures (i.e., coordinate sets) are grouped into free-energy minima, not according to geometric characteristics, but rather according to the dynamics. The network analysis shows that escape from traps with a mixture of loop and helical content is rate limiting and that there are multiple parallel folding channels originating from a kinetically partitioned unfolded state (Fig. 4). The kinetic grouping analysis, using secondary structure strings (39), reveals that, of the four main folding channels, one starts from within the helical basin, whereas the remaining three start from the following metastable states: two C-terminal α-helical turns formed (Fig. 4*B*), only N-terminal α-helical turn formed (Fig. 4*C*), and only one C-terminal α-helical turn formed (Fig. 4*D*). The unfolded state is kinetically partitioned so that the traps are connected to the one folded state in a star-like manner, and the folded basin acts as a hub (22) for the interchange between misfolded or partially folded states. An example is shown in Fig. 4*D* where, starting from the trap with only the C-terminal turn formed, the peptide first reaches the fully α-helical state and then unfolds to a structure with broken N-terminal turn (as seen from the decreasing brightness of the yellow coloring of the nodes along this pathway). Importantly, different channels have barriers of different heights (Table S1), with rates ranging from ≈1/(10 ns) for folding from the helical basin (green in Fig. 4*A*) to ≈1/(1,000 ns) for folding from the structure with only N-terminal turn formed (cyan in Fig. 4*A*). Therefore, the spread observed in the site-specific rates originates from the heterogeneous degree of formation of individual helical hydrogen bonds in different traps.

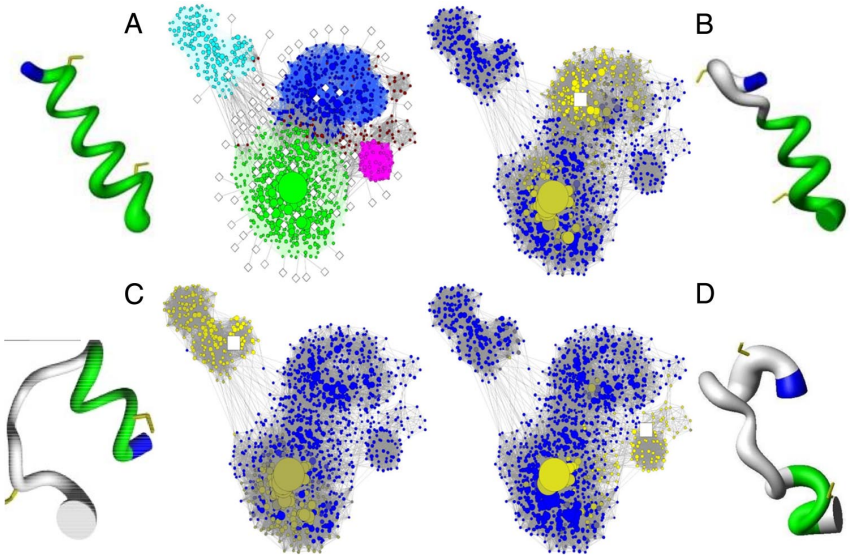


Fig. 4. The network analysis (22, 29) of the 100 MD runs at 281 K shows parallel folding channels. Each node (i.e., conformation) of the network represents a secondary structure string (39), and a link is a direct transition (within 20 ps) observed in the MD runs. The surface of each node is proportional to its statistical weight, and only the 1,387 nodes with at least 200 snapshots (≈96.7% of total sampling) are shown to avoid overcrowding. (*A*) The free-energy basins, i.e., native (green nodes) and metastable states [identified by kinetic grouping analysis (29) using a commitment time of 10 ns to group conformations that interconvert rapidly], are shown with different colors, and their characteristics are listed in Table S1. Within each basin, nodes and intrabasin links are shown with the same color, and interbasin links are colored in gray. White diamonds indicate the starting points of 82 of the 100 folding runs, whereas the remaining 18 runs reached directly the most populated node (i.e., fully formed α-helix, large green circle) and are not shown. An enlarged version of the network is shown in Fig. S2. (*B–D*) Nodes are colored according to the values of the mean first passage time (29) from the most populated node (white squares) of individual metastable states to all other nodes. The time scale goes from 0 (yellow) to >2 μs (blue). The coloring shows that the unfolded state is kinetically partitioned, and the folded (i.e., fully α-helical) state acts as a hub. Visits to unfolded metastable states different from the starting one require a much longer time than reaching the folded state. Representative structures of the folded state and each metastable state are shown by flexible tubes of variable diameter reflecting conformational disorder, with α-helical turns in green, loop segments in gray, N terminus in blue, and cysteine side chains in yellow for emphasizing the position of the linker.

The interpretation of the experimental data, thanks to the folding mechanism and pathways extracted from the MD simulations, provides strong evidence that the rate-limiting step of helix folding (at low temperature) is exactly the escape rate from a few metastable states (some of them stabilized by nonnative contacts) rather than the commonly assumed nucleation step.

Conclusion

The kinetics and mechanism of folding of a photoswitchable cross-linked α-helix have been investigated by a combined experimental–simulation study. At low temperatures, the hydrogen bond formation rates of different sites scatter by almost one order of magnitude, whereas they approach each other at higher temperatures. The spread of rates is significantly larger than the 10% effect observed for an isolated helix (12) and appears to be too large to be consistent with conventional nucleation-propagation models along the lines of ref. 16. Furthermore, with group L7, we directly observe a nonnative contact in the misfolded ensemble, the existence of which, however, is neglected in nucleation-propagation models. On the other hand, good agreement with an all-atom MD simulation is obtained, which justifies drawing interpretations at atomic resolution from the MD results. Notably, the MD analysis unmasks discrete traps (i.e., nonnative free-energy basins) along parallel folding pathways, which render the overall kinetics nonexponential. The cross-linker actually stabilizes these traps, as observed from the difference in response in MD runs with and without cross-linker (Fig. 3*F*). However, in contrast to previous studies, where noncooperative folding has been interpreted as barrierless folding (40, 41), we argue here that a completely different scenario, i.e., a few traps in the unfolded state that are separated from the native basin by barriers of different heights, may lead to a similar noncooperative behavior. This scenario (with barriers of different heights) is different from a barrierless landscape, but they share a higher population of conformations with intermediate compactness with respect to the two-state behavior (42).

Isolated α-helices fold in a marginally cooperative manner, as seen by the somewhat steeper melting curve (34, 35) than in Fig. 1*C* and the only small variation of folding rates along the peptide chain (12). If α-helix folding were cooperative, one could think of it as binary (all or nothing) when studying the folding of tertiary structures in larger proteins. However, addition of an azo-cross-linker as a switchable structural constraint finally destroys the already weak cooperativity of isolated α-helices. Structural constraints of this sort might also exist for a helical segment in a larger protein by the very fact that the helix is connected through the backbone to the rest of the polypeptide chain, and its flexibility is restrained by tertiary contacts with other pieces of the protein. As such, the folding of secondary and tertiary structures cannot be thought of as decoupled. Paradoxically, the folding of the cross-linked α-helix might be closer to the natural situation inside a globular protein than that of an isolated helix.

Materials and Methods

Experimental. Synthesis of the molecule, Ac-AACAK^{5}AAAAK^{10}AAACK^{15}A-NH_2, cross linked by a azo-moiety acting as a photoswitch and with ^{13}C = ^{18}O-labeled alanine at eight different positions, was performed as described (18). IR pump-probe spectroscopy with delays ranging from 10 ps to 40 μs was performed by using two electronically synchronized Ti:S laser systems, one of which was frequency doubled to generate pump pulses at 420 nm and the second pumped an IR-OPA to obtain broadband IR probe pulses (19, 20).

Computational. MD simulations were performed by using the CHARMM program package (31) using standard procedures and an implicit solvent (30). The force

field parameters for the azo-moiety were derived from the PARAM19 for the amide backbone and phenyl ring of Phe as well as from ref. 43 for the dihedral angles of the central N = N bond (Fig. S1). An equilibrium ensemble with the cross-linker in the *cis* conformation was sampled by two runs of replica exchange MD (44). After instantaneously switching the torsional potential of the central N = N bond to one that strongly favors the trans-configuration, ensembles of nonequilibrium Langevin dynamics runs of 4 μs each were started from the *cis* equilibrium ensemble at both 330 K and 281 K. Network analysis (22, 29) of the resulting nonequilibrium trajectories was supported by the program WORDOM (45). Because of the implicit solvent model, both the absolute temperatures and rates are somewhat arbitrary. In the comparison with the experiments, we therefore focus on trends, rather than the absolute values.

For a more detailed account of materials and methods, see *SI Text*.

ACKNOWLEDGMENTS. We thank Riccardo Pellarin for suggesting the control runs without cross-linker, Gianluca Interlandi for help in the initial set-up of the MD simulations, Rolf Pfister for the synthesis of the molecules, Francesco Rao for interesting discussions, Ben Schuler for instructive discussions and for the access to the HPLC-equipment, Bernhard Spingler and Philipp Antoni for the access to the CD equipment, and Jan Helbing and Andrea Prunotto for technical assistance. The MD simulations were run on the Matterhorn cluster of the University of Zürich. This work was supported by Swiss National Science Foundation Grants 200020-107492/1 (to P.H.) and 205320-118214 (to A.C.) and by a fellowship of the ''Forschungskredit'' of the University of Zürich (to E.H.G.B.).

1. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood? *Nat Struct Biol* 6:351–358.
2. Creighton TE (1993) *Proteins.* (Freeman, New York).
3. Zimm BH, Bragg JK (1959) Theory of the phase transition between helix and random coil in polypeptide chains. *J Chem Phys* 31:526–535.
4. Lifson S, Roig A (1961) On the theory of helix–coil transition in polypeptides. *J Chem Phys* 34:1963–1974.
5. Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL (1991) Parameters of helix–coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* 31:1463–1470.
6. Muñoz V, Serrano L (1995) Helix design, prediction and stability. *Curr Opin Biotechnol* 6:382–386.
7. Doig AJ (2002) Recent advances in helix–coil theory. *Biophys Chem* 101: 281–293.
8. Scheraga HA, Vila JA, Ripoll DR (2002) Helix–coil transitions re-visited. *Biophys Chem* 101–102:255–265.
9. Kaya H, Chan HS (2000) Polymer principles of protein calorimetric two-state cooperativity. *Proteins* 40:637–661.
10. Thompson PA, Eaton WA, Hofrichter J (1997) Laser temperature jump study of the helix–coil kinetics of an alanine peptide interpreted with a ''kinetic zipper'' model. *Biochemistry* 36:9200–9210.
11. Thompson PA, *et al.* (2000) The helix–coil kinetics of a heteropeptide. *J Phys Chem B* 104:378–389.
12. Huang C-Y, *et al.* (2002) Helix formation via conformation diffusion search. *Proc Natl Acad Sci USA* 99:2788–2793.
13. Sadqi M, Fushman D, Muñoz V (2006) Atom-by-atom analysis of global downhill protein folding. *Nature* 442:317–321.
14. Krimm S, Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv Protein Chem* 38:181–364.
15. Silva RAGD, Kubelka J, Bour P, Decatur SM, Keiderling T (2002) A site-specific conformational determination in thermal unfolding studies of helical peptides using vibrational circular dichroism with isotopic substitution. *Proc Natl Acad Sci USA* 97:8318–8323.
16. Doshi U, Muñoz V (2004) The principles of α-helix formation: Explaining complex kinetics with nucleation-elongation theory. *J Phys Chem B* 108:8497–8506.
17. Brewer SH, Song B, Raleigh DP, Dyer RB (2007) Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. *Biochemistry* 46:3279–3285.
18. Kumita JR, Smart OS, Woolley GA (2000) Photo-control of helix content in a short peptide. *Proc Natl Acad Sci USA* 97:3803–3808.
19. Bredenbeck J, Helbing J, Kumita JR, Woolley GA, Hamm P (2005) α-Helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time resolved IR spectroscopy. *Proc Natl Acad Sci USA* 102:2379–2384.
20. Ihalainen JA, *et al.* (2007) Folding and unfolding of a photoswitchable peptide. *Proc Natl Acad Sci USA* 104:5383–5388.
21. Hamm P, Helbing J, Bredenbeck J (2006) Stretched versus compressed exponential kinetics in α-helix folding. *Chem Phys* 323:54–65.
22. Rao F, Caflisch A (2004) The protein folding network. *J Mol Biol* 342:299–306.
23. Krivov SV, Karplus M (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc Natl Acad Sci USA* 101:14766–14770.
24. Hummer G, Garcia AE, Garde S (2001) Helix nucleation kinetics from molecular simulations in explicit solvent. *Proteins* 42:77–84.
25. Caflisch A (2006) Network and graph analyses of folding free energy surfaces. *Curr Opin Struct Biol* 16:71–78.
26. Chekmarev SF, Krivov SV, Karplus M (2006) Folding of ubiquitin: A simple model describes the strange kinetics. *J Phys Chem B* 110:8865–8869.
27. Chowdhury S, Zhang W, Wu C, Xiong G, Duan Y (2003) Breaking non-native hydrophobic clusters is the rate limiting step in the folding of an alanine-based peptide. *Biopolymers* 68:63–75.
28. Makowska J, *et al.* (2006) Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc Natl Acad Sci USA* 103:1744–1749.
29. Muff S, Caflisch A (2008) Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein. *Proteins Struct Funct Bioinf* 70:1185–1195.
30. Ferrara P, Apostolakis J, Caflisch A (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins Struct Funct Bioinf* 46:24–33.
31. Brooks BR, *et al.* (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
32. Nguyen PH, Stock G (2006) Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. *Chem Phys* 323:36–44.
33. Nguyen PH, Gorbunov RD, Stock G (2006) Photoinduced conformational dynamics of a photoswitchable peptide: A nonequilibrium molecular dynamics simulation study. *Biophys J* 91:1224–1234.
34. Marqusee S, Robbins VH, Baldwin RL (1989) Unusually stable helix formation in short alaninen-based peptides. *Proc Natl Acad Sci USA* 86:5286–5290.
35. Huang C-Y, Klemke JW, Getahun Z, DeGrado WF, Gai F (2001) Temperature-dependent helix–coil transition of an alanine based peptide. *J Am Chem Soc* 123:9235–9238.
36. Werner JH, Dyer RB, Fesinmeyer RM, Andersen NH (2002) Dynamics of the primary processes of protein folding: Helix nucleation. *J Phys Chem B* 106:487–494.
37. Gooding EA, *et al.* (2005) The effects of individual amino acids on the fast folding dynamics of α-helical peptides. *Chem Commun* 5985–5987.
38. Rohl CA, Baldwin RL (1998) Deciphering rules of helix stability in peptides. *Methods Enzymol* 295:1–26.
39. Andersen CAF, Palmer AG, Brunak S, Rost B (2002) Continuum secondary structure captures protein flexibility. *Structure (London)* 10:174–184.
40. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) Experimental identification of downhill protein folding. *Science* 298:2191–2195.
41. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423:193–197.
42. Knott M, Chan HS (2006) Criteria for downhill protein folding: Calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins* 65:373–391.
43. Carstens, H (2004) Conformation dynamics of light-stable peptide: Molecular dynamics simulations and data-driven model-building. PhD thesis (Ludwig Maxmilians University, Munich).
44. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
45. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A (2007) WORDOM: A program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23:2625–2627.

Supporting Information

Ihalainen *et al.* 10.1073/pnas.0712099105

SI Text

Molecular System. Alanine 1-^{13}C and $H_2{}^{18}O$ were purchased from Cambridge Isotopes Laboratories. The ^{16}O to ^{18}O exchange in the Ala-carbonyl was performed as described in ref. 1, and the FMOC-protection for the ^{13}C = ^{18}O-labeled alanine was performed as described in ref. 2. IR, together with mass spectroscopic analysis, confirmed that the sample contained >90% FMOC-Ala-1-^{13}C = ^{18}O-^{18}OH. Ac-AACAK 5AAAAK^{10}AAACK^{15}A-NH_2 labeled at eight different positions were prepared by using FMOC-based solid-phase peptide synthesis (GL Biochem). The peptides were then cross-linked with the photoisomerizable linker (3) to obtain the photoswitchable peptide. After removing TFA, the sample was diluted in D_2O to a concentration of ≈1 mM.

Experimental Setup. For the time-resolved measurements, two electronically synchronized femtosecond laser systems were used (4). The output of laser system 1 was frequency doubled to generate pulses at 420 nm, which switches the azobenzene from the *cis* to the *trans* state. The output of laser system 2 was used to pump an OPA to obtain IR probe pulses (100 fs, center frequency 1,620 cm^{-1}, bandwidth 240 cm^{-1} FWHM). The probe and reference beams were frequency-dispersed in a spectrometer and imaged onto a 2 × 32-pixel HgCdTe detector array. The sample was switched back to the initial *cis*-state by using a cw-Argon-ion laser at 366 nm, (and a properly filtered Hg-lamp for the stationary FTIR data).

Force-Field and Implicit Solvation Model. All simulations and most of the analysis of the trajectories were performed with the program CHARMM (5); the rest of the analysis was done with the program WORDOM (6), which is particularly efficient in handling large sets of trajectories. The peptide Ac-AACAR 5AAAAR^{10}AAACR^{15}A-NH_2 was simulated with and without the linker. The MD simulations were performed (at slightly higher temperature values of 8°C instead of 6°C and 57°C instead of 45°C) to compare with a previous IR spectroscopy study of the Arg-peptide (7) before the experimental work on the isotope-labeled Lys-peptide had started. The Arg/Lys difference is not expected to affect the agreement between experimental and simulation results, because of the similar side-chain length and positively charged end groups. Also, similar folding kinetics are observed for the Arg- and Lys-peptides by IR spectroscopy analysis (7). All heavy storms were considered explicitely as well as the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field). The default cutoff of 7.5 Å was used for the nonbonding interactions. A mean field approximation based on the solvent-accessible surface area was used to describe the main effects of the aqueous solvent (8).

Parametrization of the Cross-Linker. The atom types for the cross-linker atoms were derived from the PARAM19 amide backbone and phenyl ring of Phe. For the double bond of the two nitrogen atoms between the rings of the photoswitch, the parameters for the dihedral angles were taken from ref. 9. The parameters for bonds, angles, and impropers were derived per analogy with the corresponding proteinaceous fragments in PARAM19.

Dihedral Function. To emulate the photoswitching process (i.e., isomerization of the C-*n* = N-C dihedral) the energy term

$$E^{N=N}_{dihedral} = 21.6(1 + cos(2\theta - 180)) + 4.36(1 + cos4\theta) \quad [1]$$

(black curve in Fig. S1) was modified by removing the minimum at $\theta = 0°$ and using a force constant of 60 kcal/mol to preserve the shape of the function from 90° to 180°

$$E^{N=N}_{dihedral,trans} = 60.0(1 + \cos\,\theta) + 4.36(1 + \cos\,2\theta) \quad [2]$$

(red curve in Fig. S1). The Langevin dynamics simulations of folding were run with the modified dihedral term starting from snapshots saved during equilibrium *cis* runs. The isomerization took place within a few picoseconds.

MD Simulations of the Peptide with Cross-Linker in the *cis* and *trans* Conformations. The equilibrium ensemble of the peptide in the *cis* conformation of the cross-linker was sampled by two replica exchange MD (REMD) (10) simulations (9 μs and 15 μs) of six replica each at temperature values of 281 K, 304 K, 330 K, 358 K, 388 K, and 420 K. Upon merging the REMD simulation segments at 281 K (totalling 24 μs), 100 equally spaced snapshots were selected as starting structures for the folding runs at low temperature (see below). The same procedure was used to select, from simulations at 330 K, the 50 starting conformations of the folding runs at high temperature. The equilibrium ensemble of the peptide in the *trans* conformation of the cross-linker was sampled by four REMD simulations (3 μs each, for a total of 12 μs) of six replica each with the same temperature values used for the cis equilibrium REMD simulations.

MD Simulations of Folding. As in the time-resolved IR experiments, in the MD simulations, the folding process is triggered by an ultrafast isomerization of the cross-linker within a few picoseconds. Therefore, the MD runs closely mimic the photo-triggered α-helix formation. Fifty (at 330 K) or 100 (at 281 K) Langevin dynamics runs of 4 μs each were started from the equilibrium *cis* conformation of the peptide previously sampled by REMD (see above). Moreover, 50 Langevin dynamics runs at each of the two temperature values were performed for the peptide without the linker. A friction coefficient of 1 ps^{-1} was used. This value is much smaller than the one of water (43 ps^{-1} at 330 K) to allow for sufficient sampling within the microsecond time scale of the simulations. A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of 2 × 10^5 snapshots for each 4-μs run. A series of 100 4-μs runs requires two months on a 100-CPU cluster (about 0.5 μs simulation time per week on a single CPU). Using explicit water simulations it would have been impossible to obtain the 0.4 milliseconds of simulation time required to sample a statistically significant number of folding transitions at low temperature, which is a necessary condition for the present analysis.

Hydrogen Bond Analysis. The number of hydrogen atoms (covalently bound to amide nitrogen or guanidinium nitrogen atoms) close to each carbonyl oxygen was calculated along the MD trajectories to compare with the experimental signal of the isotopically labeled peptide. Oxygen–hydrogen distance cutoffs of 2.6 Å or 2.8 Å (with or without O···H–N angle cutoff of 120°) yielded essentially identical kinetic traces. Therefore, the distance cutoff of 2.6 Å (without angle criterion) was used.

1. Murphy RC, Clay KL (1990) Preparation of labeled molecules by exchange with oxygen-18 water. *Methods Enzymol* 193:338–348.
2. Bolin DR, Sytwu II, Humiec F, Meienhofer JJ (1989) Preparation of oligomer-free N-alpha-FMOC and N-alpha-urethane amino acids. *Int J Peptide Protein Res* 33:353-359.
3. Kumita JR, Smart OS, Woolley GA (2000) Photo-control of helix content in a short peptide. *Proc Natl Acad Sci USA* 97:3803–3808.
4. Bredenbeck J, Helbing J, Hamm P (2004) Continuous scanning from picoseconds to microseconds in time resolved linear and nonlinear spectroscopy. *Rev Sci Instrum* 75:4462-4467.
5. Brooks BR, *et al.*(1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
6. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A (2007) Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23:2625–2627.
7. Ihalainen JA, *et al.* (2007) Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc Natl Acad Sci USA* 104:5383–5388.
8. Ferrara P, Apostolakis J, Caflisch A (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins Struct Funct Bioinf* 46:24–33.
9. Carstens H (2004) Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modellbildung. PhD thesis (Ludwig Maxmilians University, Munich).
10. Sugita V, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141-151.
11. Muff S, Caflisch A (2008) Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein. *Proteins Struct Funct Bioinf* 70:1185–1195.
12. Andersen CAF, Palmer AG, Brunak S, Rost B (2002) Continuum secondary structure captures protein flexibility. *Structure (London)* 10:174–184.

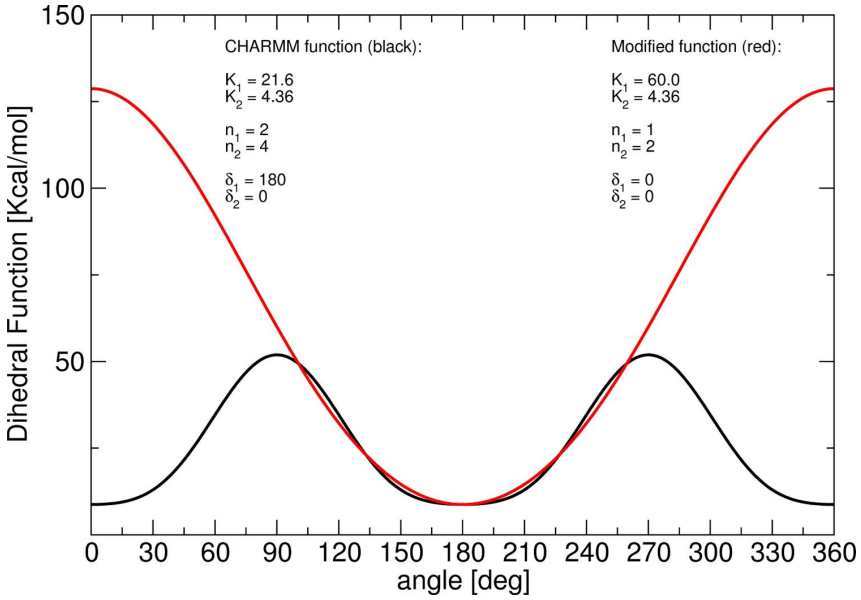


Fig. S1. Dihedral energy function: $E^{N=N}_{dihedral} = K_1(1 + \cos(n_1\theta - \delta_1)) + K_2(1 = \cos(n_2\theta - \delta_2))$.

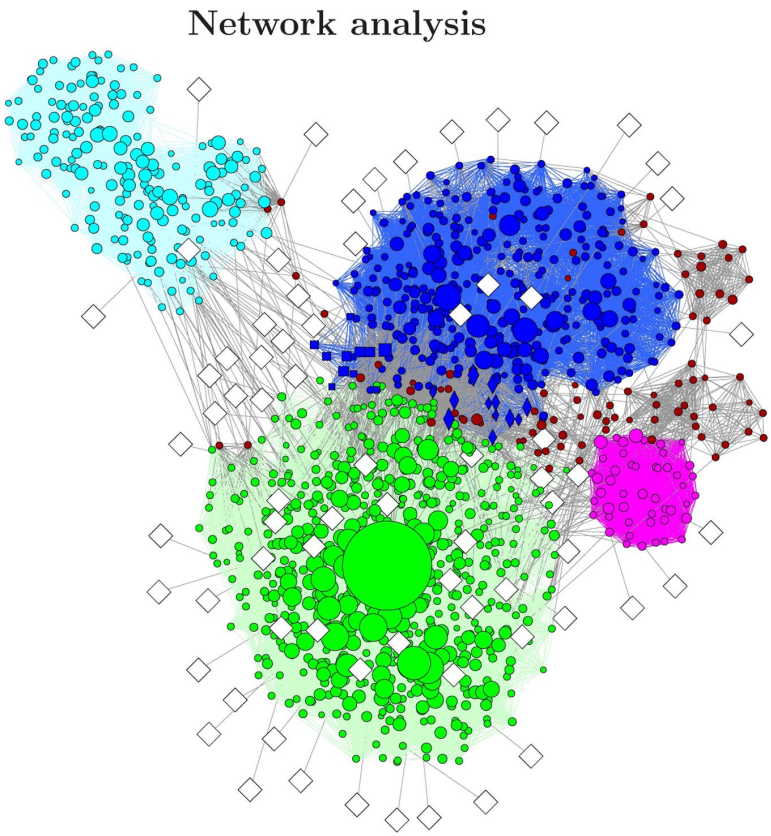


Fig. S2. See legend of Fig. 4 in main text and Table S1.

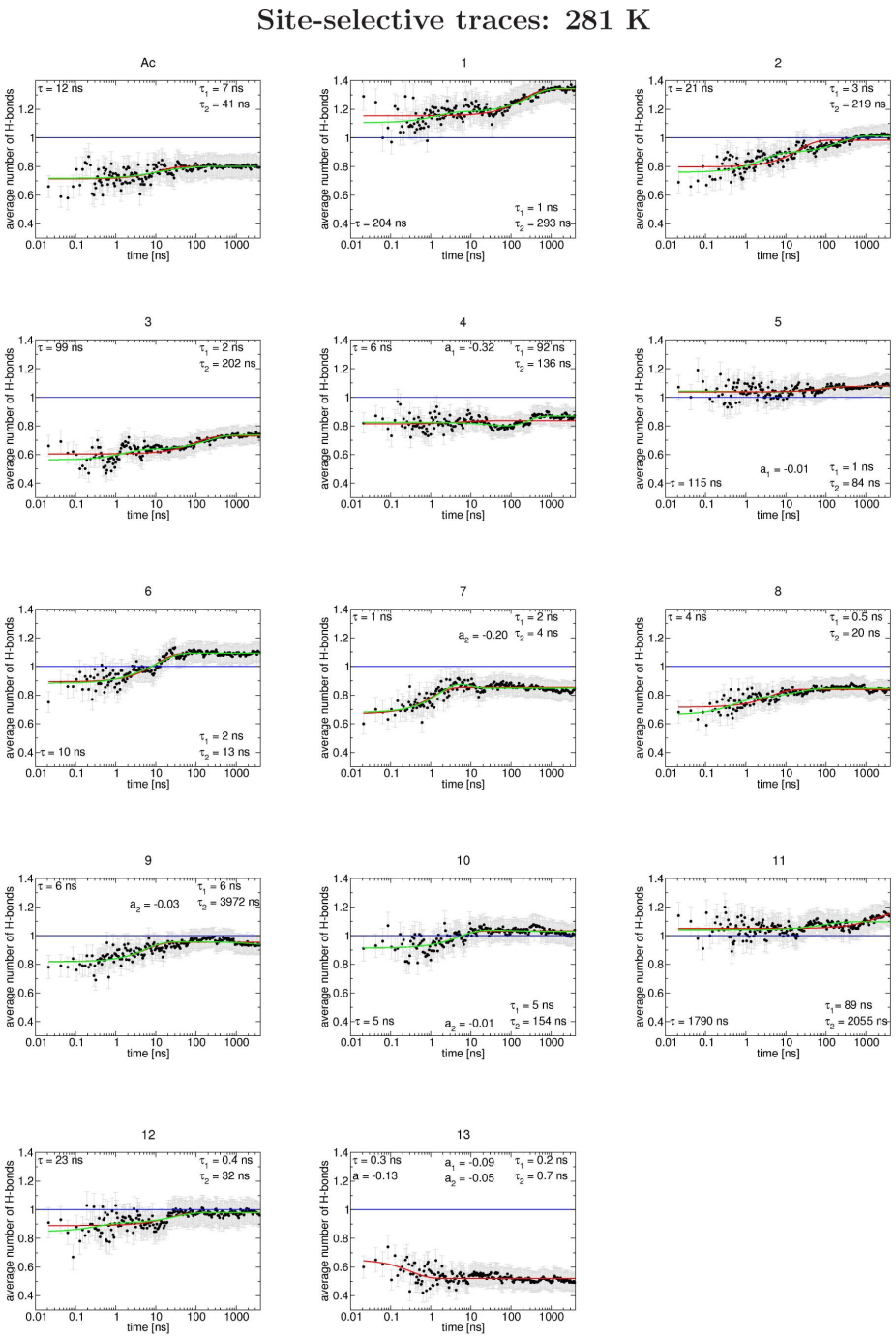


Fig. S3. Folding signal of individual carbonyl groups from Langevin dynamics simulations (first set, 281 K; second set, 330 K). The semilog plots show the time dependence of the average number of hydrogen atoms within 2.6 Å of the backbone carbonyl oxygen. Single- and double- (only at 281 K) exponential fitting curves (red and green, respectively) are shown with folding times and negative amplitudes given within each plot. All of the other values are shown in Tables S2 and S3. The blue horizontal line at $y = 1.0$ is drawn as a reference. Carbonyl groups in residues 14–16 are almost never involved in hydrogen bonds and are therefore not shown.

Site-selective traces: 330 K

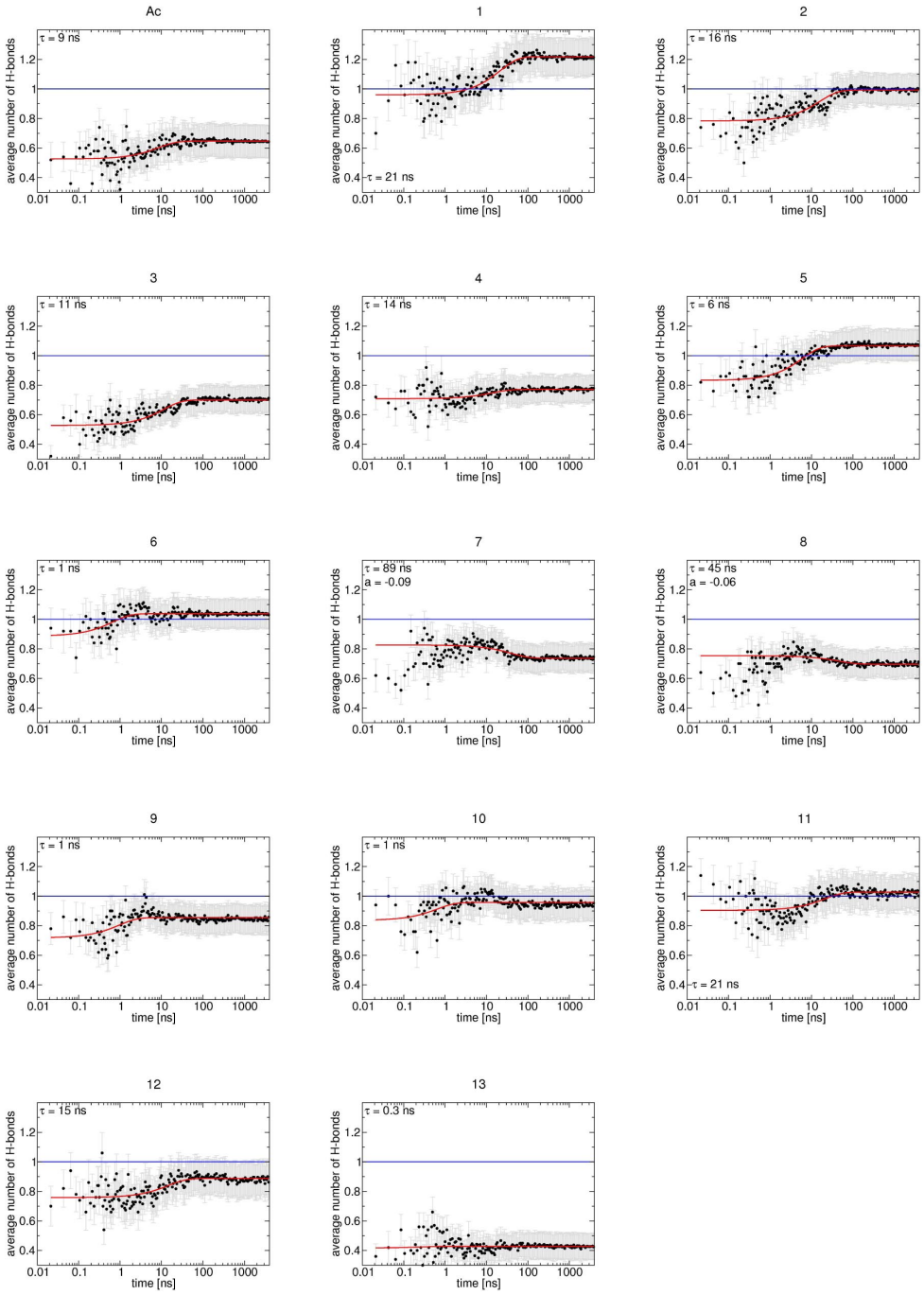


Fig. S3. Continued.

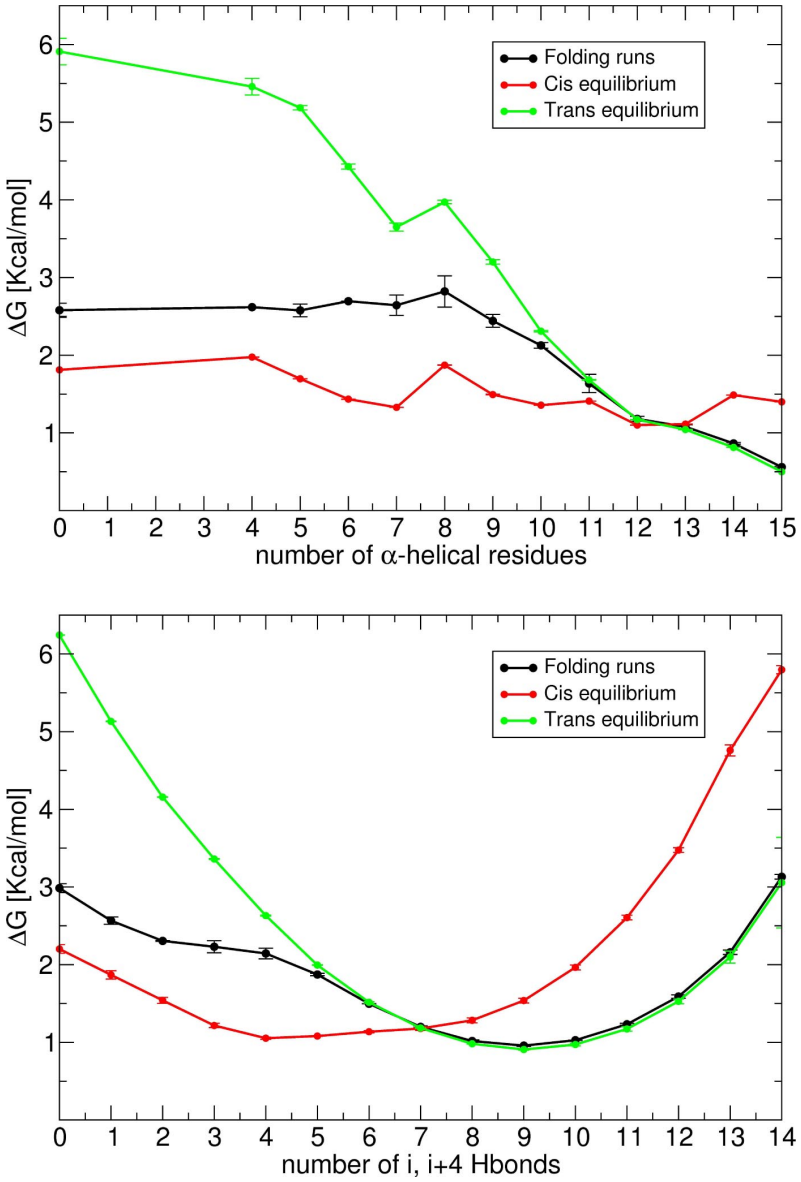


Fig. S4. Free-energy profiles at 281 K as a function of the number of α-helical residues (*Upper*) or number of α-helical backbone hydrogen bonds (*Lower*). A total of 4×10^6 conformations were used for the folding runs, 1.2×10^6 conformations for the *cis* equilibrium simulations, and 0.6×10^6 conformations for the *trans* equilibrium simulations. Values at equilibrium were extracted from REMD segments at 281 K: 24 μs for *cis* and 12 μs for *trans*. The number of α-helical residues was calculated by DSSP (12), whereas the number of α-helical hydrogen bonds was calculated by CHARMM using a cutoff of 2.6 Å for the distance between hydrogen and oxygen atoms. Error bars are calculated by block averages, where the dataset is divided into two equal subsets. In contrast to the network analysis, the complexity of the folding process is hidden in these projections along a one-dimensional progress variable.

Table S1. Free-energy basins (i.e. folded state and most populated metastable states) determined by kinetic grouping analysis (KGA) (11) for the folding runs at 281 K with τ_{commit} = 10 ns

Most populated node [conformation]	Weight, %	τ_f, ns	Helical content in basins		Color of KGA basin
			H	H + G + I	
-HHHHHHHHHHHHHHH-	28.3	5	13.6 ± 3.1	14.6 ± 3.2	Green
–TT–IHHHHHHHHH-	32.6	396	7.3 ± 2.8	8.9 ± 3.1	Blue circles
-BSSSB-HHHHHHHHH-	1.3	424	7.7 ± 2.7	8.7 ± 2.5	Blue squares
–TTS-SHHHHHHHHH-	0.1	395	7.1 ± 1.7	8.8 ± 1.5	Blue diamonds
–HHHHHII-S-TTS–	17.4	928	5.5 ± 2.4	7.9 ± 2.2	Cyan
–SSSSS–SHHHHHH-	7.0	398	4.2 ± 2.7	5.4 ± 2.9	Magenta
Others	7.0	–	–	–	Brown

Mean folding times (τ_f) are average values for snapshots in a basin. The colors indicated in the last column are those used in Fig. S2. The first column contains the string of secondary structure of native or metastable states (i.e., of KGA basins). There are eight possible ″letters″ in the secondary structure ″alphabet:″ ″H,″ ″G,″ ″I,″ ″E,″ ″B,″ ″T,″ ″S,″ and ″–″, standing for α-helix, 3_{10} helix, π-helix, extended, isolated β-bridge, hydrogen bonded turn, bend, and unstructured, respectively (12). Because the N- and C-terminal residues are always assigned an ″–″ (12), a 16-residue peptide can, in principle, assume $8^{16} \simeq 10^{14}$ conformations.

Table S2. Single- and double-exponential fit parameters for individual folding signals at 281 K from the Langevin dynamics runs

	Single exponential		Double exponential			
Residue	a	τ, ns	a_1	τ_1, ns	a_2	τ_2, ns
Ac		12	0.05	7	0.04	41
1	0.19	204	0.08	1	0.16	293
2	0.19	21	0.13	3	0.12	219
3	0.13	99	0.07	2	0.10	202
4	0.02	6	**−0.32**	92	0.36	136
5	0.04	115	**−0.01**	1	0.04	84
6	0.20	10	0.05	2	0.16	13
7	0.18	1	0.40	2	**−0.20**	4
8	0.13	4	0.11	0.5	0.07	20
9	0.14	6	0.14	6	**−0.03**	3,972
10	0.12	5	0.13	5	**−0.01**	154
11	0.11	1,790	0.05	89	0.01	2,055
12	0.09	23	0.06	0.4	0.07	32
13	**−0.13**	0.3	**−0.09**	0.2	**−0.05**	0.7
14	**−0.04**	9	-	-	-	-
15	**−0.04**	11	-	-	-	-
16	**−0.06**	2,650	-	-	-	-

Negative amplitudes are shown in bold type.

Table S3. Single-exponential fit parameters for individual folding signals at 330 K from the Langevin dynamics runs

	Single exponential	
Residue	*a*	τ, ns
Ac	0.12	9
1	0.26	21
2	0.21	16
3	0.17	11
4	0.06	14
5	0.23	6
6	0.15	1
7	**−0.09**	89
8	**−0.06**	45
9	0.14	1
10	0.12	1
11	0.12	21
12	0.13	15
13	0.01	0.3
14	**−0.11**	2
15	**−0.06**	4
16	**−0.03**	5

Negative amplitudes are shown in bold type.

Chapter 3

Bulky side chains and non-native salt bridges slow down the folding of a cross-linked helical peptide: A combined molecular dynamics and time-resolved infrared spectroscopy study.

[*JPC B*, **2009**, *113*, **4435-4442**]

Bulky Side Chains and Non-native Salt Bridges Slow down the Folding of a Cross-Linked Helical Peptide: A Combined Molecular Dynamics and Time-Resolved Infrared Spectroscopy Study

Beatrice Paoli,[†] Michele Seeber,[‡] Ellen H. G. Backus,[§] Janne A. Ihalainen,[§] Peter Hamm,[§] and Amedeo Caflisch*,[†]

Department of Biochemistry and Department of Physical Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, and Dulbecco Telethon Institute and Department of Chemistry, University of Modena and Reggio Emilia, Via Campi 183, I-41100 Modena, Italy

Received: November 27, 2008; Revised Manuscript Received: January 14, 2009

Multiple 4-μs molecular dynamics (MD) simulations are used to study the folding process of the cross-linked α-helical peptide Ac-EACAR^{5}EAAAR^{10}EAACR^{15}Q-NH_2 (EAAAR peptide). The folding kinetics are single exponential at 330 K, while they are complex at 281 K with a clear deviation from single-exponential behavior, in agreement with time-resolved infrared (IR) spectroscopy measurements. Network analysis of the conformation space sampled by the MD simulations reveals four main folding channels which start from conformations with partially formed helical structure and non-native salt-bridges in a kinetically partitioned unfolded state. The independent folding pathways explain the comparable quality of models based on stretched exponential and multiexponential fitting of the kinetic traces at low temperature. The rearrangement of bulky side chains, and in particular their reorientation with respect to the cross-linker, makes the EAAAR peptide a slower folder at 281 K than a similar peptide devoid of the three glutamate side chains. On the basis of this simulation result, extracted from a total MD sampling of 1.0 ms, a mutant with additional bulky side chains (three methionines replacing alanines at positions 2, 7, and 12) is suggested to fold slower than the EAAAR peptide. This prediction is confirmed by time-resolved IR spectroscopy.

I. Introduction

Protein folding from the ensemble of denatured conformations to the native state is a complex transition because of the many degrees of freedom and interaction centers involved.[1] In fact, folding is driven by a delicate balance of van der Waals and electrostatic forces between atoms in the protein, and protein−solvent interactions. To tackle the complexity of the folding process synergistic combinations of experimental and computational studies have been used successfully, in particular to characterize the transition state ensemble of two-state proteins,[2] and novel techniques are being developed. On the experimental side, small and ultrafast folding proteins have been discovered and further engineered.[3,4] Recently, time-resolved IR spectroscopic techniques have been used to monitor folding kinetics of helical peptides with a cross-linker from the picosecond to microsecond time range.[5] For triggering helical folding, the conformation of the cross-linker (azobenzene) is "switched" from cis to trans by using light of a given wavelength.[6] On the computational side, approaches based on complex network[7] and graph analyses[8] have been recently introduced for analyzing long MD simulations, which become more and more affordable because of the ever-increasing computer power. These graph-theoretical approaches have played an important role in showing that the surprisingly simple two-state picture of protein folding, often obtained by projecting the free energy onto an arbitrarily chosen progress variable, is not consistent with the complexity of the actual free-energy surface.[9]

Here, we investigate by MD the folding kinetics of the cross-linked peptide Ac-EACAR^{5}EAAAR^{10}EAACR^{15}Q-NH_2 (abbreviated hereafter as EAAAR), and compare the simulation results with the available time-resolved IR data.[5] In a previous work we had analyzed the kinetics of the Ac-AACAR^5-AAAAR^{10}AAACR^{15}A-NH_2 peptide (abbreviated as AAAAR).[10] One major difference between AAAAR and EAAAR is that only the latter can form salt bridges, which in the fully α-helical state involve the three pairs of side chains Glu_n-Arg_{n+4} (n = 1, 6, and 11). Therefore, we focus here on the influence of the sequence on the folding process. To also investigate the effect of the temperature, MD simulations of EAAAR folding are performed at 281 K (50 runs of 4 μs each) and at 330 K (50 runs of 0.5 μs each). The runs are started from pre-equilibrated cis ensembles at each temperature. The folding process upon ultrafast switching is emulated in the MD simulations by sudden switch of the dihedral potential of the N=N bond in the middle of the photoswitchable cross-linker. Two recently developed analysis techniques are applied to shed light on the kinetics of EAAAR folding and analyze its free-energy surface: The network of the conformation space (CS) and kinetic grouping analysis. The former is a very useful tool for the analysis and illustration of the actual (unprojected) free-energy surface and its dynamic connectivity[7,9,11] while kinetic grouping analysis is a recently introduced approach for identifying free-energy basins in long MD simulations.[12] Nonequilibrium MD simulations of an 8-residue cyclized peptide immersed in dimethylsulfoxide have been published recently[13] but the focus of that work was on the contributions to the frequency shift, while pathways and kinetics of folding were not investigated.

The main motivation of the present MD study was to provide the atomic-resolution picture needed to elucidate the effects

* To whom correspondence should be addressed. E-mail: caflisch@bioc.uzh.ch. Tel.: +41 44 635 55 21. Fax: +41 44 635 68 62.
[†] Department of Biochemistry, University of Zurich.
[‡] University of Modena and Reggio Emilia.
[§] Department of Physical Chemistry, University of Zurich.

10.1021/jp810431s CCC: $40.75
Published on Web 03/03/2009

of the sequence on the folding kinetics of α-helical peptides, which has recently emerged from the time-resolved IR measurements. In this respect, the following three questions are particularly relevant: How do bulky side chains and (non-native) salt bridges influence the kinetics of α-helix formation? Can the simulation results be used to derive a simplified but quantitative description of the free-energy surface and its sequence dependence? Is it possible to use the atomic-resolution picture obtained by the MD study to predict mutants with different folding rates? The present work shows that the network analysis of the MD simulations provide a description of the free-energy surface of α-helix formation that not only captures the complex folding behavior and sequence dependence but can also be used to make suggestions for guiding experiments. In fact, a mutant of EAAAR with three methionine side chains (replacing three alanines) is predicted to fold slower on the basis of the MD simulation results, and this prediction is validated a posteriori by additional time-resolved IR measurements.

II. Methods

A. Simulation Methods and Protocols. ***Force Field and Implicit Solvation Model.*** All simulations and most of the analysis of the trajectories were performed with the program CHARMM;[14,15] the rest of the analysis was done with the program WORDOM,[16] which is particularly efficient in handling large sets of trajectories. All heavy atoms were considered explicitely as well as the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field). For the nonbonding interactions the default cutoff of 7.5 Å was used to be consistent with the parameters of the nonbonding energy terms of the force field which were determined using this cutoff value. A mean field approximation based on the solvent accessible surface area was used to describe the main effects of the aqueous solvent.[17]

Parametrization of the Cross-Linker. The atom types for the cross-linker atoms were derived from the PARAM19 amide backbone and phenyl ring of Phe. For the double-bond of the two nitrogen atoms between the rings of the photoswitch the parameters for the dihedral angles were taken from ref 18. The parameters for bonds, angles, and impropers were derived per analogy with the corresponding proteinaceous fragments in PARAM19. Before starting the production runs, short MD simulations of the cross-linker alone were analyzed visually and showed reasonable qualitative behavior, for example, limited out-of-plane fluctuations of the *N*-acetylaniline moieties.

Dihedral Function. To emulate the photoswitching process (i.e., isomerization of the C−N=N−C dihedral) the energy term

$$E^{\mathrm{N=N}}_{\mathrm{dihedral}} = 21.6(1 + \cos(2\theta - 180)) + 4.36(1 + \cos 4\theta) \quad (1)$$

(black curve in Figure 1) was modified by removing the minimum at $\theta = 0°$ and using a force constant of 60 kcal/mol to preserve the shape of the function from 90° to 180°

$$E^{\mathrm{N=N}}_{\mathrm{dihedral,trans}} = 60.0(1 + \cos\theta) + 4.36(1 + \cos 2\theta) \quad (2)$$

(red curve in Figure 1). The Langevin dynamics simulations of folding were run with the modified dihedral term starting from snapshots saved during equilibrium cis runs. The isomerization took place within a few picoseconds.

REMD Simulations of the Peptide with Cross-Linker in the Cis Conformation. The equilibrium ensemble of the peptide in the cis conformation of the cross-linker was sampled by a replica exchange MD (REMD)[19] simulation of six replica at temperature values of 281, 304, 330, 358, 388, and 420 K. Temperature exchange attempts were performed every 20 ps (10000 MD steps) as in previous implicit solvent REMD simulations of helical and extended peptides,[10,20] and the acceptance ratio ranged between 0.27 and 0.33. Each replica was sampled for 18 and 27 μs for EAAAR and EMAAR, respectively. From the 18 μs (27 μs) of REMD sampling at 281 K, 50 (100) snapshots saved along time intervals of constant length were selected as starting structures for the EAAAR (EMAAR) folding runs at low temperature (see later and Table 1). The same procedure was used to select, from simulations at 330 K, the 50 and 130 starting conformations of the folding runs at high temperature of EAAAR and EMAAR, respectively.

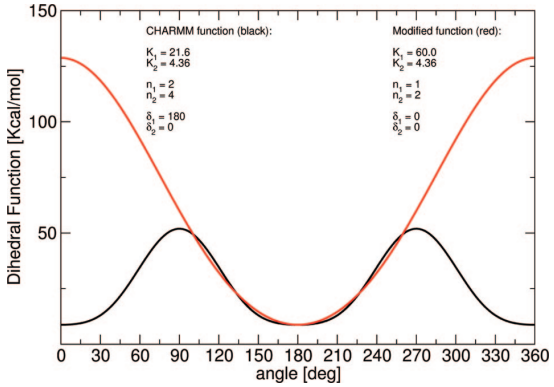


Figure 1. Emulation of the photoswitching by change in the dihedral energy term of the N=N bond. The dihedral energy term is $E^{\mathrm{N=N}}_{\mathrm{dihedral}} = K_1(1 + \cos(n_1\theta - \delta_1)) + K_2(1 + \cos(n_2\theta - \delta_2))$.

MD Simulations of Folding. As in the time-resolved IR experiments, in the MD simulations the folding process is triggered by an ultrafast isomerization of the cross-linker within a few picoseconds. Therefore, the MD runs closely mimic the phototriggered α-helix formation. In a previous study we simulated the folding of the AAAAR peptide.[10] The same protocol is used here for the folding runs of EAAAR and EMAAR. At 281 K, 50 Langevin dynamics runs of 4 μs each were started from the equilibrium cis conformation of the EAAAR peptide previously sampled by REMD (see previous and Table 1). For the EMAAR peptide 100 runs of 8 μs each were performed at 281 K, and 130 runs of 0.5 μs each at 330 K. A friction coefficient of 1 ps^{-1} was used in all simulations. This value is much smaller than the one of water (43 ps^{-1} at 330 K) to allow for sufficient sampling within the μs time scale of the simulations. A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of 2×10^5 snapshots for each 4-μs run. A series of 100 4-μs runs requires two months on a 100-CPU cluster (about 0.5 μs simulation time per week on a single Opteron 2.4 GHz CPU). Using explicit water simulations it would have been impossible to obtain the 1.0 ms of MD sampling (50 4-μs runs of EAAAR and 100 8-μs runs of EMAAR) required to collect a statistically significant number of folding transitions at low temperature, which is a necessary condition for the present analysis.

Coarse-Graining. There are several ways for assigning snapshots (i.e., coordinate sets) to coarse-grained conformations (nodes and strings are used as synonyms in this paper) and different types of analysis might require different coarse-graining approaches.[7,21] The coarse-graining used in this work is based on secondary structure strings.[22] A conformation is a single string of secondary structure, for example, the most populated

TABLE 1: Simulations Performed

	equilibrium cis state sampled by REMD		folding runs[a] 281K		330K		
	no. runs	length[b] [μs]	no. runs	length [μs]	no. runs	length [μs]	ref
AAAAR	2	24	100	4	50	0.5	10
EAAAR	1	18	50	4	50	0.5	this work
EMAAR	3	27	100	8	130	0.5	this work

[a] An ensemble of snapshots, saved at time intervals of constant length along the REMD simulation segments at a given temperature, was selected as starting structures for the folding runs at that temperature. [b] These values refer to individual replicas and not to the cumulative time which is six times larger.

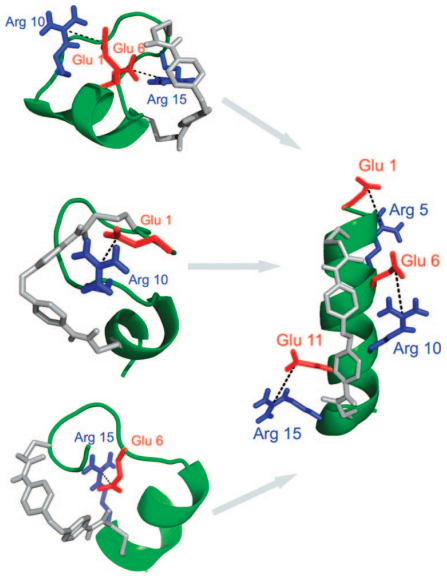


Figure 2. Folding of the cross-linked α-helical peptide EAAAR involves rearrangements of charged side chains. (Left) Three snapshots extracted from free-energy basins of the cis equilibrium simulations show non-native salt bridges. (Right) The most populated conformation with the cross-linker in the trans state is fully helical, and the three native, i.e., n, $n + 4$, salt bridges are populated at about 50%.

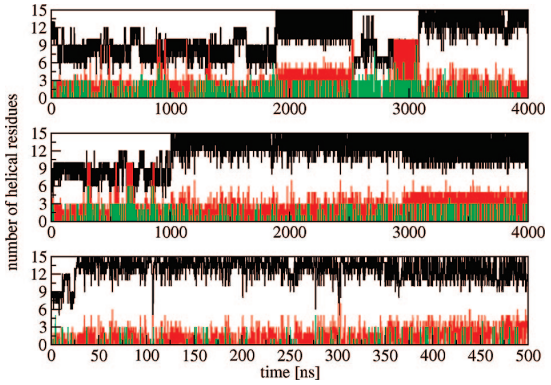
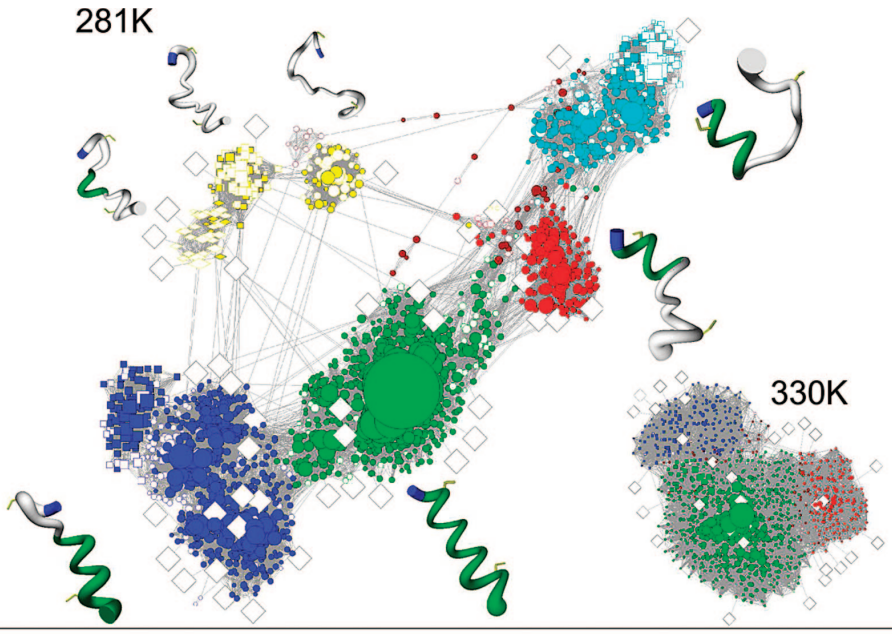


Figure 3. Increase of helical content in the MD simulations of the EAAAR peptide upon switching from the cis to the trans term of the dihedral energy of the N=N bond in the cross-linker. Time series of representative simulations are shown at 281 K (top and middle) and 330 K (bottom). The number of residues in stretches of α-helix, π-helix, and 3_{10}-helix is in black, red, and green, respectively. The bottom time series shows only the first 500 ns to better resolve the folding event in the first 25 ns.

conformation of EAAAR with trans cross-linker is $-H_{15}-$. There are eight possible "letters" in the secondary structure "alphabet": "H", "G", "I", "E", "B", "T", "S", and "−", standing for α helix, 3_{10} helix, π helix, extended, isolated β-bridge, hydrogen-bonded turn, bend, and unstructured, respectively. Since the N- and C-terminal residues are always assigned an "−",[22] an N-acetylated 16-residue peptide can in principle assume $8^{15} \simeq 10^{13}$ conformations. Note that the vast majority of these conformations are sterically forbidden. The secondary structure-based coarse-graining is appropriate for structured peptides without loops and has three advantages with respect to approaches based on root-mean-square deviation (rmsd) of atomic coordinates. First, it does not require the use of an arbitrarily chosen threshold value. Second, each node is uniquely defined by its secondary structure string which is a useful conformational "label". Third, the same type of secondary structure classification[22] is used for coarse-graining and analysis of structural properties (e.g., helical content in Figure 3).

CS Network. Conformations (i.e., secondary structure strings) are nodes of the CS network and the direct transitions between them are links.[7] Given the number $\tilde{w}$ of snapshots with a unique secondary structure string, the statistical weight w of a node is equal to $w = \tilde{w}/N$, where N is the total number of snapshots sampled in the folding runs. At 281 K, the 20 ps saving frequency and 50 4-μs folding runs yield $N = 10^7$ for EAAAR. The CS networks are illustrated using only nodes with $\tilde{w} \geq 50$ to avoid overcrowding. Two nodes (in the CS network of "heavy" nodes) are connected by a link if they either include a pair of snapshots that are visited within 20 ps or they are separated by one or more nodes with less than 50 snapshots each. It is important to underline that the CS network of "heavy" nodes is used solely for illustrative purposes, whereas all of the quantitative analysis in this work, including kinetic grouping and fitting of average number of helical residues, was performed using the complete trajectory, that is, all snapshots and all nodes.

Kinetic Grouping Analysis. In long MD runs the complete information about the native and non-native free energy basins is present in the trajectory. Therefore, peptide conformations can be grouped into free energy minima according to rapid transitions at equilibrium. This approach, called kinetic grouping analysis, has been developed previously and used to analyze the differences in the free energy surface of a 20-residue three-stranded β-sheet peptide and its single-point mutant Trp10Val,[12] both simulated at folding−unfolding equilibrium. Kinetic grouping analysis is based on the observation that if two conformations interconvert rapidly they are not separated by a barrier and therefore belong to the same basin. The method requires only one parameter, the commitment time τ_{commit}, which is a typical relaxation time within the basins of the system. Different values of τ_{commit} allow one to analyze different levels of ruggedness of the free energy surface.[12] Values of τ_{commit} ranging from 1 to 20 ns were tested in this study. The final analysis was performed with $\tau_{commit} = 10$ ns and $\tau_{commit} = 1$ ns at 281 and 330 K, respectively, because these values allow for intrabasin relaxation and are much shorter than the time required for interbasin



Most populated node in individual basins	Percentage of unfolded	τ_f (ns) Node	Basin	Helical content in basins H	H+G+I	color
-HHHHHHHHHHHHHHH-	–	–	6	13.5 ± 1.7	14.7 ± 1.7	green
-HHHHHHIII-IIIII-	15	84	93	8.0 ± 2.8	13.0 ± 1.7	red
-BSSB-HHHHHHHHHH-	49	843	601	8.0 ± 2.0	9.5 ± 1.3	blue circles
--TT--HHHHHHH-TT-	5	389	710	6.9 ± 0.6	7.0 ± 1.2	blue squares
-HHHHHHII-S-TTS--	17	1772	1480	5.8 ± 1.2	7.6 ± 1.1	cyan circles
-IIHHHHHI-S-TTS--	6	3292	3260	4.2 ± 1.5	7.5 ± 1.7	cyan squares
-TTT-S-TT-IIIII-	4	922	840	1.9 ± 2.6	6.1 ± 1.9	yellow circles
--S--GGGII-IIIII-	3	540	525	2.9 ± 3.1	8.1 ± 1.6	yellow squares
-BSSB-HHHH-IIIII-	1	273	313	5.3 ± 3.2	8.1 ± 2.1	yellow diamonds

Figure 4. The CS network of the folding runs of EAAAR at 281 K shows multiple channels to the folded state (green nodes). Each node (i.e., conformation) of the network represents a secondary structure string and a link is a direct transition observed in the MD runs. The surface of each node is proportional to its statistical weight and only the 2218 nodes with at least 50 snapshots (98.7% of the total sampling) are shown to avoid overcrowding. Conformations sampled only in the folding runs and not in the trans equilibrium simulation are shown by white nodes with colored rims. White diamonds indicate the starting points of 45 of the 50 folding runs while the remaining 5 runs reached directly the most populated node and are not shown. The free-energy basins identified by kinetic grouping analysis[12] are shown with different colors or symbols, and their characteristics are listed in the table where folding times (τ_f) are average values for snapshots in the most populated node of individual basins (column Node) or the entire basin (column Basin). Brown nodes were not assigned by kinetic grouping analysis. Representative structures of individual basins are shown by flexible tubes of variable diameter reflecting conformational disorder, with α-helical segments in green, N-terminus in blue, and cysteine side chains in yellow for emphasizing the position of the cross-linker. The CS network of the folding runs at 330 K is also shown (bottom right). This figure was made using visone (www.visone.de) and MOLMOL.[27]

transitions. The probability $p_{commit}(i \rightarrow j)$ to observe a transition from node i to node j within a given τ_{commit} is an asymmetric, directed measure of the *kinetic similarity* of nodes i and j. Once the p_{commit}-matrix has been calculated for highly populated nodes (the 500 heaviest nodes in this work), pairs of nodes (i, j) are grouped together if $p_{commit}(i \rightarrow j) \geq 0.5$. Lighter nodes are then assigned in a postprocessing step.[12] This way of grouping leads to a partitioning into disjoint sets, that is, a disconnected CS network, whose subgraphs correspond to different basins. Although the kinetic grouping analysis has been introduced to analyze long equilibrium simulations, it can be also applied to nonequilibrium simulations. In fact, since kinetic grouping analysis takes into account only the rapid, local interconversion between conformations within a basin, there is no requirement for global equilibrium sampling. Therefore, also basins visited during kinetic simulations can be isolated, provided that the system equilibrates locally.

B. Experimental Methods. The experimental setup for the time-resolved measurements has been described previously.[5,10,23]

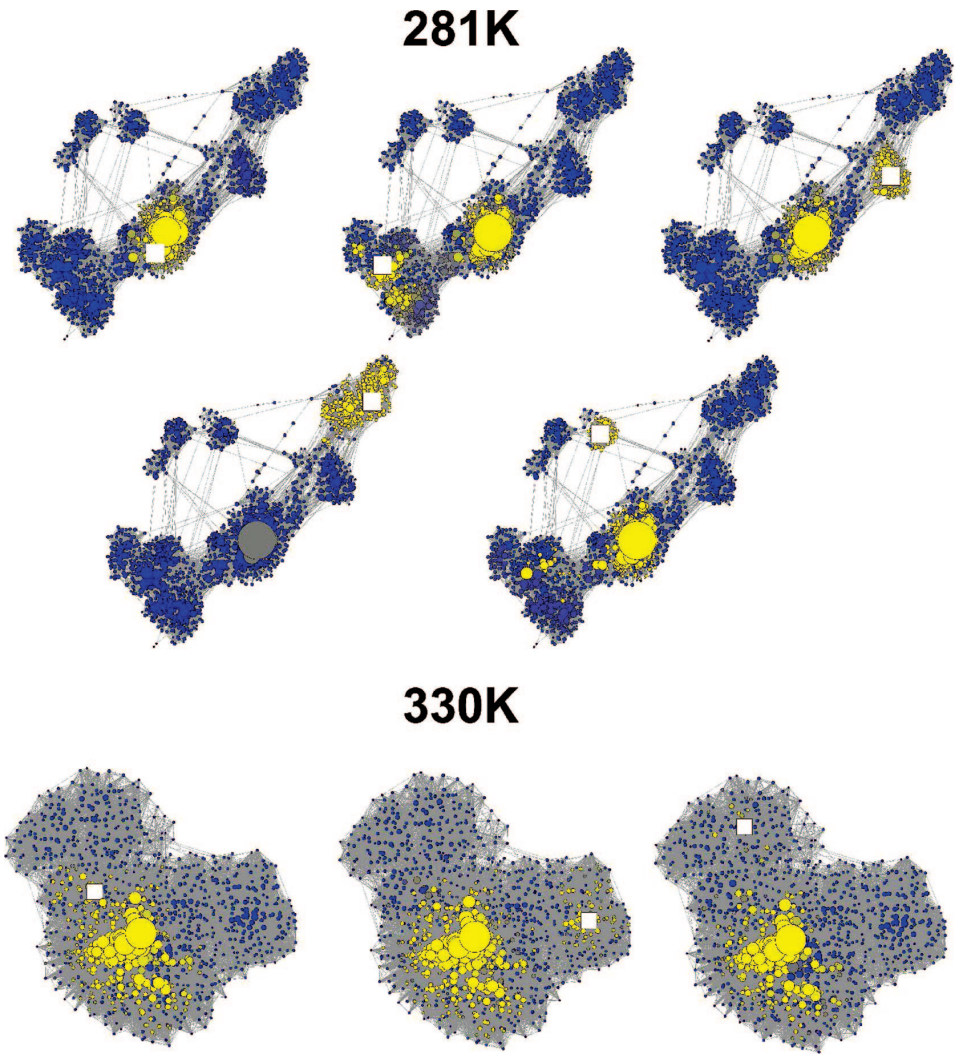


Figure 5. The unfolded state of the EAAAR peptide is kinetically partitioned. Nodes are colored according to mean first passage times[12] from the most populated node (white square) of individual free-energy basins to all other nodes in the CS network. (Top and middle) Folding runs at 281 K. The time scale ranges from 0 (yellow) to 4 μs (blue) except for the top, left network where the range is 0−1 μs. The 13th most populated node of the native basin was used as "starting node" in the top, left network. Most nodes within the basin of the starting node are visited relatively fast (yellow), indicating rapid intrabasin transitions and supporting the results of the kinetic grouping analysis (Figure 4). Visits to unfolded basins different from the starting one is much slower (blue) than reaching the folded state (olive or yellow) which shows that the unfolded state is kinetically partitioned. In other words, the folded state is a hub.[7,12,23] (Bottom) Folding runs at 330 K. The time scale ranges from 0 (yellow) to 0.1 μs (blue).

Briefly, two electronically synchronized femtosecond laser systems were used. The output of the one laser system was frequency doubled to generate pump pulses at 420 nm which switches the azobenzene from the cis to the trans state. The output of the second laser system was used to pump an OPA to obtain IR probe pulses (100 fs, center frequency 1620 cm^{-1}, bandwidth 240 cm^{-1} fwhm), that could continously be delayed with respect to the pump-pulses from about 10 ps to 40 μs. The probe pulses were frequency dispersed in a spectrometer and imaged onto a 32 pixel HgCdTe detector array. A "folding signal" was generated from the amide I band, which frequency-shifts upon the strengthening of the backbone hydrogen bonds in the helical state. The sample was circulated in a closed-cycle flow cell optimized for small sample volumes. To prepare a well-defined initial condition, unfolded peptides with the azobenzene moiety in the cis-conformation were produced by constantly illuminating the sample with an excess of light at 366 nm, generated either by a cw-Argon-ion laser or a high-power LED.

III. Results and Discussion

The equilibrium cis and trans ensembles of the EAAAR peptide are presented before the folding runs of which they represent the initial and final states, respectively (Table 1).

A. Unfolded State (Cross-Linker in Cis Conformation) of the EAAAR Peptide. In the equilibrium cis simulations, the helicity (i.e., sum of residues in α-, 3_{10}- or π-helical conformation as measured by the program DSSP)[22] of the EAAAR peptide is about 65% at both temperature values. Different conformations with partial helical content are observed but predominant structures are not present. At 281 K, several free energy basins are identified by kinetic grouping analysis and

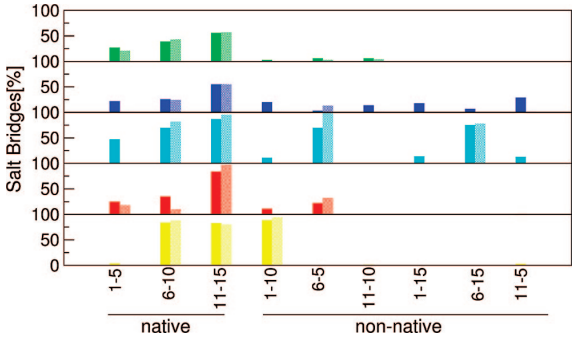


Figure 6. Salt bridge frequencies in individual free-energy basins of folding runs of the EAAAR peptide at 281 K. Values for entire basin and most populated node are displayed by filled and hatched bars, respectively. Colors correspond to those used in Figure 4.

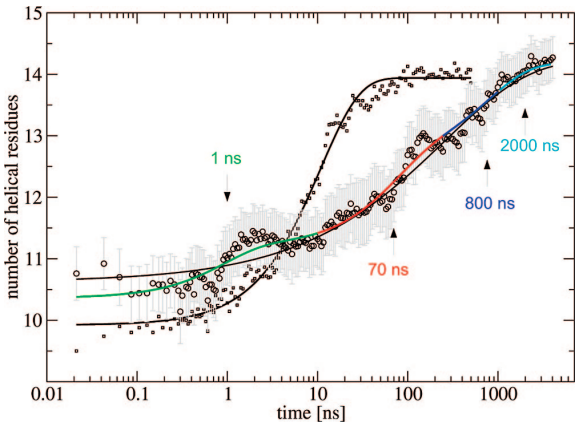


Figure 7. Kinetic traces of folding of the EAAAR peptide from MD simulations at 281 (circles) and 330 K (squares). The four-exponential fitting curve at 281 K [14.2 − 0.8 exp(−*x*/1) − 1.4 exp(−*x*/67) − 1.4 exp(−*x*/755) − 0.2 exp(−*x*/1770), $\chi^2 = 3.3$] is colored according to the folding "channels" identified by kinetic grouping and CS network analyses, i.e., as in Figure 4, and arrows denote the four time constants. The stretched-exponential fitting of the data at 281 K [14.3 − 3.9 exp(−*x*/185)$^{0.40}$, $\chi^2 = 3.5$], and the single-exponential fitting of the data at 330 K [13.9 − 4.0 exp(−*x*/11), $\chi^2 = 1.9$] are shown by black solid lines. Error bars are not shown for the data points at 330 K to avoid overcrowding and because the single-exponential fitting curve is always within the error bars.

only three of them have a statistical weight larger than 10%. The most populated basin (containing about one-quarter of the snapshots) has an unstructured N-terminal segment and two α-helical turns at the C-terminal segment. Individual basins have homogeneous distributions of helical content and salt bridges. In several basins non-native salt bridges are more frequent than the three salt bridges between residues Glu_n and Arg_{n+4} (with $n = 1$, 6, and 11) termed *native* salt bridges hereafter (Figure 2).

B. Helical State (Cross-Linker in Trans Conformation) of the EAAAR Peptide. In the equilibrium trans simulations, the helicity of the EAAAR peptide is about 90% and 80% at 281 and 330 K, respectively. These values can be compared with the helicity measured by CD: 93% and 64% at 281 and 322 K, respectively.[5] The agreement is good if one takes into account the 10% error of the CD measurements. On the other hand, the influence of the temperature is less pronounced in the MD simulations which is due mainly to the lack of explicit temperature-dependence in the solvation model. Therefore, in the comparison with the experiments we focus on qualitative trends rather than the quantitative comparison of absolute values.

The fully α-helical structure (secondary structure string $-H_{15}-$) is the most populated conformation (i.e., node) in the trans state of the EAAAR peptide with a statistical weight of 38% and 29% at 281 and 330 K, respectively. The corresponding free-energy basin has a population of about 90% at 281 K. In the fully α-helical state, the three native salt bridges are formed about 25−55% and 20−45% of the time at 281 and 330 K, respectively, while non-native salt bridges are sporadic (Supporting Information). Therefore, the main differences between the equilibrium cis and trans ensembles are that the former has a significantly less pronounced helical content and a much higher frequency of non-native salt bridges.

C. Folding Kinetics and Complexity of Free-Energy Surface of the EAAAR Peptide. At each temperature value, 50 snapshots saved with constant frequency along the corresponding equilibrium cis simulation were used as starting conformations for the kinetic runs (see Methods and Table 1). The time series of the helical content during the 281 K runs show that the folding transition takes place after time intervals of different length (Figure 3). Furthermore, folding at 330 K is much faster than at 281 K.

CS Network Analysis Reveals Kinetic Partitioning. The network representation of the folding runs at 281 K shows the presence of multiple folding channels and time scales (Figure 4). Seventeen of the 50 folding runs start from the free-energy basin with unfolded N-terminal segment (blue region in Figure 4). The slowest folding channel starts from a free-energy basin with α-helical N-terminal segment and residues 7−16 unstructured (cyan basin). It is useful to color the CS network of the folding runs according to values of mean first passage time (mpft) from individual free-energy basins (Figure 5, top). Folding from the non-native basins proceeds directly to the fully helical state without passing through other basins. The barriers between individual free energy basins in the unfolded state and the fully helical state are lower than barriers within the unfolded state. The observation of kinetic partitioning of the unfolded state by kinetic grouping analysis provides further evidence to the fact that the native basin of EAAAR acts as a hub.[7,12,23] Such centrality of the native state has a significant influence on the folding kinetics and their complexity at low temperature (see also subsection Kinetic Models and Fitting of Helical Traces . These simulation results are consistent with recent experimental reports on competing folding routes and kinetic partitioning in ubiquitin[24] and the DNA-binding domain of p53.[25] Importantly, the present results provide evidence that folding pathways depend crucially on the unfolded state, that is, the starting ensemble. Therefore, in vitro folding experiments under strongly unfolding conditions (e.g., high concentration of chemical denaturants) do not necessarily give insights into the folding process under physiological conditions.

The CS network at 330 K consists of only the blue, red, and green (i.e., folded) basins (inset of Figure 4). Moreover, the unfolded state is kinetically partitioned (Figure 5, bottom), and the folding barriers are small and of similar height. The significant differences in folding kinetics at 281 and 330 K are indicative of the complexity of the system and the role of the entropy. Note that the implicit solvation model used in the MD simulations does not have an explicit temperature dependence so that only qualitative conclusions on temperature effects are possible.

Salt Bridges. As mentioned above, the EAAAR peptide has Glu and Arg side chains at positions n and $n + 4$, respectively, along the sequence. Strikingly, the stability of non-native salt bridges correlates with the length of the characteristic folding

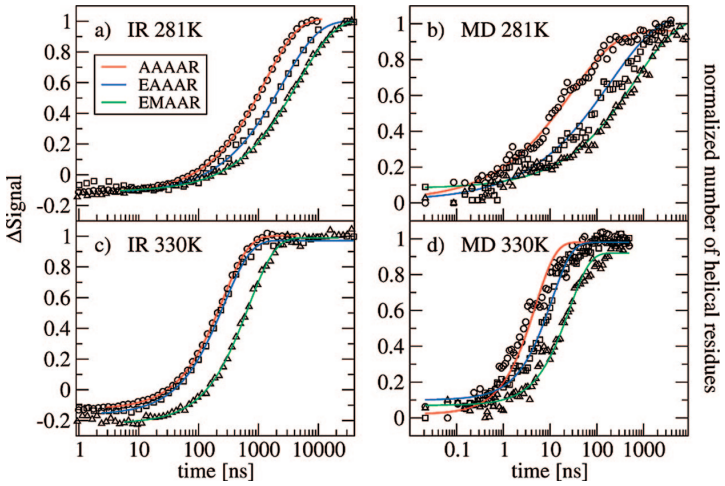


Figure 8. Comparison of normalized kinetic traces measured by time-resolved IR spectroscopy (a and c) and helical content, i.e., normalized number of residues in α, π, or 3_{10} conformation, along the MD simulations (b and d). Circles, squares, and triangles are used for AAAAR, EAAAR, and EMAAR, respectively. The fitting curve at 281 and 330 K is a stretched exponential (a and b) and a single exponential (c and d), respectively. The stretching factors at 281 K are in the range 0.4−0.5 and 0.7−0.8 for the MD and IR data, respectively. This discrepancy originates from the overestimation of the intrinsic barriers in the MD simulations, as previously observed in the large spread of folding rates of individual residues.[10] This overestimation is due in part to the predominance of intrasolute friction in the Langevin dynamics with implicit solvent and small friction coefficient (see Methods). Experimental data for AAAAR and EAAAR (a and c) are adapted from refs 5, 23.

times of the different channels (compare Figure 6 with Figure 4). In particular, the slowest channel, that is, folding from the cyan basin, is characterized by the stability of the non-native salt bridges Glu_6-Arg_5 and Glu_6-Arg_{15}, and two other slow channels by the presence of the salt bridges Glu_1-Arg_{10} (yellow basin) and $Glu_{11}-Arg_5$ (blue basin). Moreover, visual analysis of some of the 281 K folding runs indicate that rearrangement of charged side chains is hindered by the presence of the atoms of the cross-linker. In this context, it is interesting to note that the Asp14Ala mutant of the five-helix bundle protein λ_{6-85} was observed experimentally to fold faster than the wild-type[4] which is probably a consequence of the involvement of the wild-type Asp14 side chain in non-native salt bridges in the denatured state.

Kinetic Models and Fitting of Helical Traces. It is interesting to monitor the average number of helical residues during the folding process and compare it to the time-resolved IR spectroscopy data. At 330 K, the MD kinetic trace can be fitted by a single exponential (Figure 7). On the other hand, single-exponential fitting is not appropriate at 281 K, and the data are best fitted by a multiple exponential fit or a stretched exponential. While the latter is in agreement with the time-resolved IR spectroscopy data,[5] the multiple exponential fit is consistent with the CS network, which shows multiple channels connecting individual unfolded basins to a hub-like helical state (Figure 4) and kinetic partitioning of the unfolded state (Figure 5).

Stretched exponential fitting is often preferred to multiexponential fitting because the former requires only two parameters (a time constant and a power of the stretching exponent) while the latter needs two parameters for each exponential term. On the other hand, if it is possible to relate the multiexponential behavior to a specific model, like the CS network of EAAAR at 281 K, then valuable insight into the folding process is obtained. Another interesting example of multiexponential fitting is the simple kinetic model used to describe the folding of ubiquitin.[26]

D. Validation of the Atomistic Picture by Prediction of a Mutant Peptide. The time series of helical content during folding of the EAAAR and AAAAR peptides can be fitted by a stretched exponential at 281 K (Figure 8b) with time constants of 200 ± 50 and 30 ± 2 ns, respectively. As shown in Figure 2, the folded (i.e., fully helical) state has a unique arrangement of the Glu and Arg side chains and their orientation with respect to the cross-linker. Since the unfolded state consists of basins with non-native salt bridges, the slower folding of the EAAAR peptide with respect to AAAAR is mainly due to rearrangement of salt bridge pairing and entanglement of bulky side chains with the cross-linker. For this reason, we thought that a mutant with additional "long" side chains would show slower kinetics of folding, and therefore we decided to investigate a peptide with three additional Met side chains, Ac-EMCAR^{5}EMAAR^{10}EMACR^{15}Q-NH_2 (EMAAR peptide). One-hundred 8-μs simulations of folding of EMAAR from the equilibrium cis ensemble at 281 K confirmed our prediction (Figure 8b). In particular, the fitting of helical content by a stretched exponential yields a time constant of 650 ± 75 ns, which is slower than those of EAAAR and AAAAR. Note that the slower folding of EMAAR is not due to interactions between Met side chains, because the π-helical arrangment, which would bring the $n, n+5$ side chain pairs in close contact, is sporadic (Supporting Information).

To obtain an experimental validation of our prediction and the MD results, the three-point mutant EMAAR was synthesized (with exactly the same cross-linker as in AAAAR and EAAAR) and its photoswitched folding was investigated by time-resolved IR spectroscopy. At 281 K the folding of EMAAR is slower than EAAAR according to the IR traces in agreement with the MD simulation results (Figure 8a,b). The folding rates obtained by stretched-exponential fitting of the IR kinetic traces at 281 K are 4600 ns, 2500 ns, and 1300 ns for EMAAR (this study), EAAAR (ref 5), and AAAAR (ref 23), respectively. Note that at high temperature the kinetics are much faster and single exponential according to both MD simulations and time-resolved IR spectroscopy traces (Figure 8c,d). Furthermore, the rank-order of the time constants is the same as at 281 K in both simulations and experiments except for the AAAAR and EAAAR peptides that have essentially identical kinetic traces at high temperature in the experiments. In conclusion, the EMAAR mutant

was designed on the basis of the MD analysis of the AAAAR and EAAAR peptides to further investigate the mechanism of folding and the role of the sequence. Notably, the time-resolved IR data on the cross-linked α-helical peptides are consistent with and can be interpreted by the detailed picture that has emerged from the atomistic simulations.

IV. Conclusions

Multiple MD simulations of the folding of the cross-linked helical peptide EAAAR have been performed to shed light on pathways and kinetics, with the aim of obtaining a detailed (i.e., atomistic) interpretation of the available time-resolved IR spectroscopy data. The cross-linker restricts the accessible conformational space. Yet, the following three observations provide strong evidence that (a large part of) the complexity of protein folding is still present. First, with respect to an unlinked peptide, folding of the cross-linked peptide might resemble more the structuring of a protein segment in the context of the remaining of the polypeptide chain.[10] Second, there is a significant dependence of the folding kinetics on the temperature in both MD simulations and time-resolved IR spectroscopy measurements. Third and most importantly, different sequences show different folding rates indicating that the primary structure has a strong influence on the free-energy surface.

Previous MD simulations have shed light on the complex kinetics of folding of the cross-linked AAAAR peptide,[10] which in contrast to EAAAR cannot form salt bridges. The multiple folding routes and kinetic partitioning of the unfolded state, which were both revealed by the CS network and kinetic grouping analyses, suggested that the kinetics of helix folding are determined by the exiting rate(s) from traps rather than nucleation of the first helical turn. At low temperature different barrier heights were observed for different folding channels, which explained the multiexponential kinetics. The role of the side chains was not investigated.

Here, the folding kinetics of the cross-linked EAAAR peptide and their temperature dependence observed in the MD runs are not only consistent with the time-resolved IR spectroscopy data[5] but provide atomic detail explanations of the complex kinetics. In particular, analysis of MD simulations unmasks the important role of bulky side chains and non-native salt bridges. The charged side chains involved in non-native contacts in the unfolded state have to rearrange because in the fully α-helical state only the three salt bridges between residues Glu_n and Arg_{n+4} are stable. This rearrangement is hindered by the presence of the cross-linker. On the basis of this observation it was predicted that the EMAAR mutant, which has three additional Met side chains, should fold even slower than the EAAAR peptide. This prediction was substantiated by a comparison of the MD simulations of EAAAR and EMAAR, and validated experimentally.

In conclusion, a detailed picture of the folding mechanism has emerged from the MD simulations of cross-linked α-helical peptides and has been used to interpret the kinetic traces measured experimentally. The combined MD simulation and time-resolved IR spectroscopy study of the three peptides indicates that non-native interactions among the charged side chains and entanglement of bulky side chains with the cross-linker are responsible for the stretched exponential kinetics at low temperature and rank order of folding rates (AAAAR faster than EAAAR faster than EMAAR). Note that the cross-linked peptides used in the present study represent the situation of a helical segment inside a protein because the cross-linker reduces the backbone flexibility, and the interactions between the side chains and the cross-linker reflect the tertiary contacts between the helical segment and other parts of the protein, respectively. Since the three peptides AAAAR, EAAAR, and EMAAR have the same cross-linker, the "tertiary contacts" have a stronger influence on the folding kinetics than the reduced backbone flexibility.

Acknowledgment. We thank Stefanie Muff for interesting discussions and comments to the manuscript. We also thank our co-workers of refs 5 and 23, Jens Bredenbeck, Jan Helbing, Rolf Pfister, and Andrew Woolley, who significantly contributed to the experimental data of AAAAR and EAAAR (Figure 8a,c). We are grateful to Armin Widmer for the program WITNOTP which was used for visual analysis of the MD trajectories. The MD simulations were performed on the Matterhorn cluster of the University of Zurich, and we gratefully acknowledge the support of C. Bolliger and A. Godknecht. This work was supported by Swiss National Science Foundation grants to A.C. and P.H.

Supporting Information Available: Formation of native and non-native salt bridges. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11.
(2) Daggett, V.; Fersht, A. R. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497.
(3) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sat, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863.
(4) Yang, W.; Gruebele, M. *Nature* **2003**, *423*, 193.
(5) Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2379.
(6) Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Proc. Natl. Acd. Sci. U.S.A.* **2000**, *97*, 3803.
(7) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299.
(8) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766.
(9) Caflisch, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71.
(10) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9588.
(11) Gfeller, D.; De Los Rios, P.; Caflisch, A.; Rao, F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817.
(12) Muff, S.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1185.
(13) Nguyen, P. H.; Gorbunov, R. D.; Stock, G. *Biophys. J.* **2006**, *91*, 1224.
(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(15) Brooks, B. R.; et al. *J. Comput. Chem.*; in press.
(16) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625.
(17) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 24.
(18) Carstens, H. Master's Thesis. LMU M374nchen, Faculty of Physics, 2004.
(19) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(20) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. *J. Chem. Phys.* **2004**, *121*, 10748.
(21) Hubner, I. A.; Deeds, E. J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17747.
(22) Andersen, C. A. F.; Palmer, A. G.; Brunak, S.; Rost, B. *Structure* **2002**, *10*, 174.
(23) Ihalainen, J. A.; Bredenbeck, J.; Pfister, R.; Helbing, J.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5383.
(24) Crespo, M. D.; Simpson, E. R.; Searle, M. S. *J. Mol. Biol.* **2006**, *360*, 1053.
(25) Butler, J. S.; Loh, S. N. *J. Mol. Biol.* **2005**, *350*, 906.
(26) Chekmarev, S. F.; Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 8865.
(27) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics Modell.* **1996**, *14*, 51.

JP810431S

Bulky side chains and non-native salt bridges slow down the folding of a cross-linked helical peptide: A combined molecular dynamics and time-resolved infrared spectroscopy study

SUPPLEMENTARY MATERIAL

B. Paoli[1], M. Seeber[2], E.H.G. Backus[3], J.A. Ihalainen[3], P. Hamm[3], and A. Caflisch[1*]

[1]*Department of Biochemistry,*
University of Zurich,
Winterthurerstrasse 190,
CH-8057 Zurich, Switzerland
tel: +41 44 635 55 21,
fax: +41 44 635 68 62,
e-mail: caflisch@bioc.uzh.ch

[2]*Dulbecco Telethon Institute and Department of Chemistry,*
University of Modena and Reggio Emilia,
Via Campi 183, I-41100 Modena, Italy

[3]*Department of Physical Chemistry,*
University of Zurich

(Dated: January 12, 2009)

Keywords: free-energy surface, stretched-exponential folding kinetics, complex network analysis, non-native interactions, molecular dynamics

Native and non-native salt bridges

	281K		330K	
	Cis	Trans	Cis	Trans
Native salt-bridges				
Glu_1-Arg_5	20	26	14	20
Glu_6-Arg_{10}	39	34	30	35
Glu_{11}-Arg_{15}	56	55	37	45
Non-native salt-bridges				
Glu_1-Arg_{10}	22.0	3.5	10.9	2.7
Glu_1-Arg_{15}	7.2	0.3	6.2	0.5
Glu_6-Arg_5	12.5	6.3	10.1	5.8
Glu_6-Arg_{15}	15.2	0.5	8.7	0.9
Glu_{11}-Arg_5	8.3	0.5	6.5	0.7
Glu_{11}-Arg_{10}	10.6	9.1	8.5	0.9

TABLE S-I: Percentages of native and non-native salt bridges during the REMD equilibrium runs of EAAAR peptide at 281 K and 330 K. The cutoff for salt-bridges formation is 6Å between atom CD of Glu and atom CZ of Arg.

Conformational space network of EAAAR at 330K

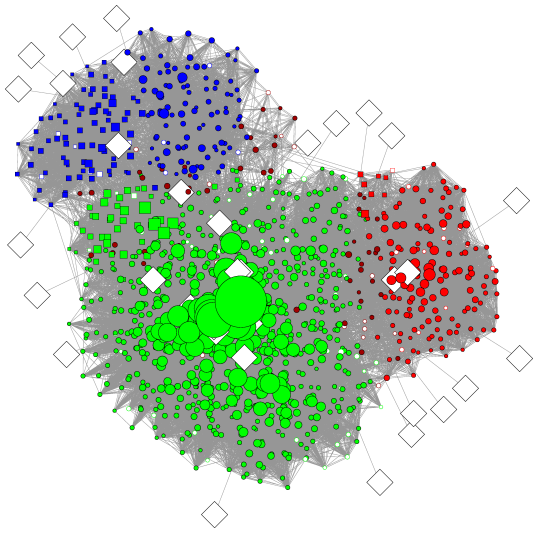


FIG. S1: The CS network of the EAAAR folding runs at 330K. The basins isolated by kinetic grouping analysis[1] with τ_{commit}=1 ns are shown in different colors as indicated in Table S-II. The surface of each node is proportional to its statistical weight and only the 1200 nodes with at least 100 snapshots (i.e. 97.7% of the total sampling) are shown to avoid overcrowding. White diamonds indicate the starting points of 46/50 folding runs while the remaining 4 runs reached directly the most populated node and are not shown. Empty circles represent nodes that are not visited during the Trans equilibrium simulation.

Most populated node in individual basins	Percentage of unfolded	τ_f (ns) Node	Basin	Helical content in basins H	H+G+I	color
-HHHHHHHHHHHHHH-	–	–	0.4	13.3 ± 2.7	14.4 ± 2.6	green circles
--SSHHHHHHHHHHH-	21	17	16	11.1 ± 2.5	11.7 ± 2.3	green squares
-HHHHHHII-IIIII-	42	10	9	8.8 ± 2.8	12.4 ± 2.7	red circles
--HHHHHT---HHHH-	1	30	26	8.2 ± 2.9	8.9 ± 3.2	red squares
--SSS-HHHHHHHHH-	21	25	18	8.1 ± 2.3	9.8 ± 1.9	blue circles
-------HHHHHHHH-	14	25	19	7.6 ± 2.2	8.8 ± 2.2	blue squares

TABLE S-II: Results of kinetic grouping analysis[1] for the folding runs with τ_{commit}=1 ns. The colors indicated in the last column are those used in Fig. S1.

Met-Met interactions in EMAAR peptide

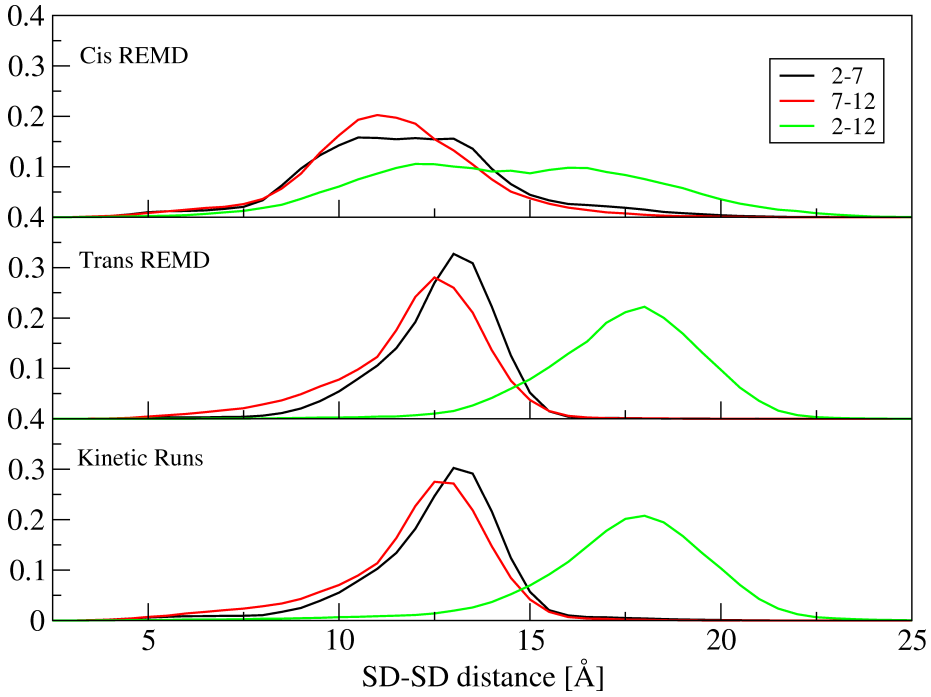


FIG. S2: Met-Met distance ditribution of EMAAR peptide in equilibrium REMD simulations (top and central panels) and kinetic runs at 281 K (bottom panel). The typical π-helical arrangment, with interactions between $n, n+5$ side chain pairs, is sporadic.

---

* corresponding author, tel: +41 44 635 55 21, fax: +41 44 635 68 62, e-mail: `caflisch@bioc.uzh.ch`

[1] S. Muff and A. Caflisch, Proteins: Structure, Function, and Bioinformatics **70**, 1185 (2008).

Chapter 4

Cross-linked helical peptides have complex free energy surface

[*to be submitted*]

Cross-linked helical peptides have complex free energy surface

B. Paoli et al.

Department of Biochemistry,

University of Zurich,

Winterthurerstrasse 190,

CH-8057 Zurich, Switzerland

tel: +41 44 635 55 21,

fax: +41 44 635 68 62,

Abstract

The formation of α-helices is one of the fundamental processes in protein folding. The folding of helical peptides can be triggered by using model peptides with a built-in photoswitchable cross-linker. Three cross-linked peptides with different amount of bulky side chains were investigated previously by experimental and computational approaches. Free energy profiles combined with root mean square deviation clustering of the MD conformations revealed that the helical ensemble of the cross-linked peptides is divided by a free-energy barrier into two regions having Arg10 on the right and left side with respect to the cross-linker. Moreover, the right to left transition (and viceversa) is hindered by the presence of the bulky cross-linker. To investigate the effect of the cross-linker we replaced it with a distance constraint mimicking the effect of a "non-bulky" cross-linker and eventually removed it. When the space previously occupied by the cross-linker is made free either by replacing or removing the cross-linker, the folding kinetics become faster and simpler compared to the cross-linked case. This indicates that the main effect of the cross-linker on the overall kinetics is due to its entanglement with the side chains rather that its effect as a distance constraining.

Keywords: molecular dynamics, distance constraint, stretched-exponential kinetics, free-energy profiles

I. INTRODUCTION

Because the helix is a common secondary structural motif in proteins, it is important to understand the folding process of the α-helix in detail as a step to understand the protein folding mechanism. A well-established way of studying physico-chemical properties of an α-helix is to use model peptides with enhanced helix-propensity[1]. In the last years, the employment of peptides with an attached azobenzene moiety acting as a photoswitchable constraint has been established as an efficient experimental tool to control the helix stability[2,3] and to trigger folding[4,5]. Two cysteines are cross-linked in such a way that the azo-moiety in the trans (cis) conformation stabilizes (destabilizes) the helix. Moreover, the cis $\rightarrow$ trans isomerization of the cross-linker is used to initiate the folding process and folding can be observed as the dominant kinetic contribution. In recent works[4–7], three peptides, whose sequences are Ac-EACAR^{5}EAAAR^{10}EAACR^{15}Q-NH_2 (hereafter noted as EAAAR), Ac-AACAR^{5}AAAAR^{10}AAACR^{15}A-NH_2 (hereafter noted as AAAAR) and, Ac-EMCAR^{5}EMAAR^{10}EMACR^{15}Q-NH_2 (hereafter noted as EMAAR) were employed to tackle the complexity of the folding process from both experimental and computational side. The combined MD simulations and time-resolved IR spectroscopy study of the three peptides[7] showed that the stretched exponential kinetics observed at 281 K and the rank order of folding rates (AAAAR faster that EAAAR faster than EMAAR) is due to the combined effect of non-native interactions among the charged side chains and entanglement of bulky side chains with the cross-linker. Nevertheless, the role of side chains was not investigated in previous works. To achieve this task, we investigated in detail the projected free-energy surface of the helical ensemble. The common way to investigate the free-energy surface is to display it as a function of a small number of order parameters as the number of native contacts[8]. The main disadvantage of this commonly used projections is the possibility of hiding essential information concerning the free-energy surface[9]. Recently, it was introduced a new progress coordinate which is able to preserve barriers and minima in the order they are met during folding/unfolding events[10,11]. It uses the normalized partition function of a given region as the progress coordinate and determines the free-energy barriers as a function of the coordinate. Root mean square deviation (RMSD) was used to coarse-grain the conformations visited during the simulations. The advantage with respect to secondary structure coarse-graining used so far[6,7] is that RMSD coarse-graining takes into account dif-

ferent side chain configurations. The result is a one-dimensional projected free-energy profile (cFEP, standing for cut-based free-energy profile, see Section IV C) which was particularly effective to describe the free-energy basins arising from different side chain configurations. To investigate the effect of the entanglement between the bulky cross-linker and the side chains, we studied the folding of the α-helical peptides replacing the cross-linker with a distance constraint which mimics the effect of a "non-bulky" cross-linker as well as without the cross-linker (free-system). It is worth to mention the novelty of our approach. In fact, the substitution of the cross-linker by a distance constraint, possible only in the context of simulations, allows us to investigate an intermediate situation in which we remove the entanglement between the cross-linker and the side chains without eliminating the action of the cross-linker as a distance constraint. This property is fundamental to control helix stability and to trigger folding and it was the reason for the design of photoswitchable peptides. Additionally, the employment of peptides with a different number of side chains was helpful to quantify the extent of the entanglement and to compare with the cross-linked case studied previously[7].

II. RESULTS AND DISCUSSION

A. mfpt-based cFEP

The equilibrium ensemble of the peptides in the cis and trans conformation of the cross-linker was sampled by REMD. An ensemble of snapshots, saved at time intervals of constant length along the cis equilibrium simulations at 281 K, was selected as starting structures for the folding runs and used for each condition (cross-linked, constrained, and free) for sake of comparison.

Visual analysis of some of the 281 K equilibrium and folding runs indicates that the rearrangement of bulky side chains (especially Arg) is hindered by the presence of the cross-linker (Fig. 1). This happens for all the sequences investigated independently from the number of bulky side chains. In particular, we observed that in the helical state (i.e., $-\mathtt{H}_{15}-$ secondary structure string[12]) the Arg in position 10 can adopt two different orientations with respect to the cross-linker: the first has Arg10 on the right side (R-conformation, Fig. 1 left panels) whereas the second has Arg10 on the left side (L-conformation, Fig. 1 right panels).

The cut-based free energy profile (cFEP, see Methods and Ref.[11]) for the helical ensemble (i.e., with the cross-linker in the trans conformation) sampled by REMD segments at 281 K is the projection of the free energy on the relative partition function Z_A/Z, a progress coordinate that increases monotonically with the distance from the reference state[10]. The side chain RMSD clusters were defined as described in Section IV C. The cutoff values chosen to distinguish the two orientations of Arg10 are 2.0Å for AAAAR and 2.5Å for EAAAR and EMAAR. For the equilibrium runs in the trans configuration of the cross-linker the number of clusters found with the aforementioned RMSD cutoffs is 2617 for AAAAR, 2366 for EAAAR, and 2513 for EMAAR. The heaviest cluster has R-configuration of Arg10, and was used as the target for cFEPs calculation. The ten most populated secondary structure strings of the heaviest cluster with R- and L-configurations are reported in Supp. Mat. (Table S-II and Table S-III). In the cFEP of each peptide we observe a free energy barrier separating R- from L-configuration (Fig. 2). The height of the barrier is about 2 kcal/mol for all peptides. Different atom selections for RMSD clustering show that Arg10 is the main responsible of the barrier. In fact, the RMSD clustering made on all side chains, only on the arginines, and only on Arg10 show similar profiles with similar barrier height, whereas the clustering made on arginines but Arg10 shows a smaller barrier (Fig. 3). Moreover, the barrier moves towards smaller values of Z_A/Z on going from AAAAR to EMAAR, indicating that the relative weight of the L-configurations increases when the number of side chains increases. To accomplish the R- to L- transition and viceversa the helical state has to partially unfold in the middle part because of the entanglement with the cross-linker (tubular structures at the top of the barriers in Fig. 2). To investigate the effect of the cross-linker on the kinetics new folding runs were performed both replacing the cross-linker with a distance constraint and eventually removing it (see Section II B). The observation of a barrier separating two states having the same secondary structure string but different side chain configurations suggests that the definition of the native state using secondary structure coarse graining is not correct. In fact, the string $\texttt{-H}_{15}\texttt{-}$ represents only a part of the native state as conformations labeled as $\texttt{-H}_{14}\texttt{--}$, $\texttt{-H}_{13}\texttt{I}_2\texttt{-}$, etc. belong also to the heaviest (i.e., native) RMSD-cluster (see Supp. Mat. Table S-II) and moreover, only $\texttt{-H}_{15}\texttt{-}$ strings with R-configuration should be considered (Fig. 4). They are separated from conformations having L-configuration by a free-energy barrier which was not detectable by using secondary structure coarse-graining.

B. Folding kinetics in the presence of a distance constrains and for the free system

The observation that different side chain configurations are hindered by the presence of the cross-linker, suggested the study of the folding process for free and for constrained systems. The latter is the main new element of the present study: by exploiting MD, we can mimic the effect of "non-bulky" cross-linker by replacing it with a distance constraint. It is interesting to monitor, for all peptides, the normalized number of helical residues during the folding process (Fig. 5). Data were fitted using a stretched-exponential function. In the presence of the cross-linker, the folding time τ strongly increases as the number of side chains increases on going from AAAAR to EAAAR to EMAAR. The stretching factor β is between 0.40 and 0.50 indicating the complexity of the system. As observed in a previous work[7], this complexity arises from the presence of multiple folding channels with different time scales originating from the kinetic partitioning of the unfolded state (i.e. the starting conformations for the kinetic runs act as kinetic traps that are rate-determining). When the cross-linker is substituted with a distance constraint, mimicking the effect of a "non-bulky" cross-linker and subsequently removed (free system), the folding kinetics becomes faster (lower τ) and simpler (higher β). The folding runs with the distance constraint represent an intermediate case in which we still observe a complex kinetics, indicated by a stretched-exponential behavior (though with a higher stretching factor ranging from 0.55 to 0.76). The rank order of the folding times observed in previous works[6,7] in the presence of the cross-linker (AAAAR faster than EAAAR faster than EMAAR) is retained also in the presence of the distance constraint, but the folding times are much smaller (about one order of magnitude) and similar to those obtained for free systems indicating that the increased number of side chains still slows down the kinetic, but to a smaller extent compared to the cross-linked case. The ratio of folding times between cross-linked and constrained peptides is 7 for AAAAR, 13 for EAAAR, and 38 for EMAAR. As we can expect, the entanglement between the cross-linker and the side chains is more pronounced for the peptide with more side chains. At the end, when the linker is totally removed, the kinetics approach, for AAAAR and EAAAR, a single-exponential behavior (as indicated by the increased stretching factors), whereas a significant stretching factor is still observed for EMAAR. For this peptide it seems that an additional effect due to self-entanglement (i.e., between side chains) should be considered. In conclusion, the remotion of the entanglement between the side chains and the cross-linker,

by substituting it with a distance constraint, makes the folding kinetics as fast as for the free-system, indicating that the cross-linker has negligible effect on the overall kinetics and its main effect is due to the interplay with the side chains rather than the constraint it represents for the peptide.

III. CONCLUSIONS

The folding kinetics of cross-linked helical peptides have been previously studied by experiments and simulations[4–7]. It has been already observed that the folding of the cross-linked peptides might resemble more the situation of a helical segment inside a protein where the cross-linker has a twofold effect: it reduces the backbone flexibility and it reflects the tertiary contacts between the helical segment and the rest of the polypeptide chain[6,7]. Here we analyzed in detail the MD ensemble of the helical conformations by projecting its free-energy surface on a new progress coordinate that preserves barriers and minima[10,11]. The resulting cFEPs showed, for each peptide investigated, a barrier arising from two different orientations of Arg10 with respect to the cross-linker. The role of other side chains is of minor importance. Moreover, the novelty of our approach consists decoupling the twofold effects of the cross-linker on backbone and side chains possible only in the context of MD simulations. In fact, non-bulky cross-linker is emulated by substituting it with a distance constraint, which mimics only the influence of the cross-linker on the backbone. An indication of the effect of the cross-linker on the kinetics of the folding process is the observation of a crossover from a slower, stretched exponential kinetics in the presence of the cross-linker towards a faster, single-exponential kinetics when the linker occupies no space, i.e., when it is replaced by a distance constraint or removed. Moreover, the longer the side chains, the slower is the kinetics, since the side chains show entanglement also among each other.

IV. METHODS

A. Molecular dynamics simulations

Force field and implicit solvation model. All simulations and most of the analysis of the trajectories were performed with the program CHARMM[13]; the rest of the analysis was done with the program WORDOM[14], which is particularly efficient in handling large sets

of trajectories. All heavy atoms were considered explicitly as well as the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field). The default cutoff of 7.5 Å was used for the nonbonding interactions. A mean field approximation based on the solvent accessible surface area was used to describe the main effects of the aqueous solvent[15]. This choice is justified by the fact that using explicit water simulations it is not possible to sample a statistically significant ensemble for the kinetic runs.

Substitution of the cross-linker with a constraint. To mimic the effect of a non-bulky cross-linker, we employed a harmonic function (i.e., an elastic interaction with no molecular counterpart) between the sulfur atoms (S) of the cysteine residues, at the same position where the cross-linker was previously attached. The distance constraint was modeled to emulate and maintain the typical S-S distance observed when the cross-linker is in trans conformation (Fig. 6, bottom panel, black curve). In this case the average S-S distance is 16.6 Å). The function is null in a given interval and harmonic at the borders (Fig. 6, top panel), as follows:

$$E(R) = \begin{cases} \frac{1}{2}K_{min}(R - R_{min})^2 & R < R_{min} \\ 0 & R_{min} < R < R_{max} \\ \frac{1}{2}K_{max}(R - R_{max})^2 & R > R_{max} \end{cases} \quad (1)$$

where $K_{min} = K_{max} = K$ are the harmonic force constants, R_{min} and R_{max} are the minimum and maximum values of the interval where the function $E(R)$ vanishes. Several tests (see Supp. Mat. Fig. S1 and Table S-I) using different K values indicated that the best agreement with the trans S-S distance distribution is obtained using R_{max}-R_{min}=1Å and K=2.5 kcal/mol Å^2 (Fig. 6, top panel, red curve).

REMD simulations of the peptide with cross-linker in the cis/trans conformation. The equilibrium ensemble of the peptide in the cis conformation of the cross-linker was sampled by a replica exchange MD (REMD)[16] simulation of six replica at temperature values of 281, 304, 330, 358, 388, and 420 K (see Table I). Upon merging the REMD simulation segments at 281 K, the same snapshots (50/100 for AAAAR and EMAAR and 50/50 for EAAAR) previously used for the folding runs of the cross-linked peptides[6,7] were selected as starting structures for the folding runs of the constrained and free systems. The equilibrium ensemble of the peptide in the trans conformation of the cross-linker was sampled by REMD simulations with the same number of replicas and temperature values used for the cis equilibrium REMD simulations (see Table I).

MD simulations of folding. Previously we simulated the folding of the cross-linked AAAAR (Ref.[6]), EAAAR (Ref.[7]), and EMAAR (Ref.[7]) peptides. The same protocol is used here for the folding runs in the presence of the constraint and for the free system. For constrained and free systems we employed the same starting snapshots of the cross-linked case. This starting ensemble is representative of the equilibrium simulations with the cross-linker in the cis conformation, which is taken as the "reference" system. The number of the performed kinetic runs and their time-length is reported in Table II. A friction coefficient of 1 ps^{-1} was used in all simulations to allow for sufficient sampling within the μs time scale of the simulations. A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of 2×10^5 snapshots for each 4-μs run.

B. Coarse-graining

A molecular dynamics trajectory is nothing but a long series of microscopic configurations visited only once. For this reason the analysis of the system needs a coarse graining of the trajectory that allows the grouping/clustering of similar snapshots/configurations. There are several procedures to efficiently achieve coarse graining and different types of analysis might require different coarse-graining approaches. For a structured peptide like α-helical peptides, root mean square deviation (RMSD) and secondary structural coarse-graining are obvious possibilities[17–19]. The coarse-graining used in this work is based on the side chains RMSD[20] and uses the leader algorithm in the version implemented by S. Krivov[11]. One advantage of RMSD clustering with respect to secondary structural coarse-graining is that side chain atoms are taken into account in the former but not in the latter. This allows us to investigate in detail the role of the entanglement of the side chains with the cross-linker on the folding process. Moreover, secondary structure-based coarse-graining has a higher tendency to group kinetically distant snapshots (i.e., snapshots in microstates separated by a high free-energy barrier[11]) because conformations having the same secondary structure string can be separated by a barrier. As an example we can consider the top panel of Fig.1, where both conformations have fully formed α-helical structure, but side chain configurations are different. According to secondary structure coarse-graining these snapshots should be grouped together, whereas according to RMSD they belong to different clusters which are separated by a free-energy barrier. In other words, the RMSD-based clustering yields less

spurious transitions[11] between microstates that are not directly connected. On the other hand, in the secondary structure-based coarse-graining each node is uniquely defined by its secondary structure string, which is a useful conformational "label". According to the DSSP program[12] each residue of a protein can be either of eight symbols `-` (coil), `E` (extended strand in a β ladder), `S` (bend), `T` (hydrogen bonded turn), `B` (residue in an isolated β-bridge), `G` (3_{10}-helix), `H` (α-helix), `I` (π-helix). The most populated structure for each studied peptide is $\texttt{-H}_{15}\texttt{-}$.

C. Cut-based Free Energy Profiles (cFEP)

A progress coordinate that preserves the barriers and minima in the order that they are met during folding/unfolding events was introduced by Krivov and Karplus[10]. It uses the relative partition function as the progress coordinate and determines the free energy barriers as a function of the coordinate by a method based on the folding probability, pfold. The procedure gives almost identical results if pfold is replaced by the mean first passage time (mfpt) to a selected node[11], which is used in the present work and briefly presented in what follows.

Given a network, the partition function of each node is given by $Z_i = \sum_i c_{ij}$, where c_{ij} is the edge capacity from node j to node i which is proportional to the number of direct transitions from j to i. When the nodes are partitioned into two groups A and B according to the minimum cut procedure[10], then $Z_A = \sum_{i\in A} Z_i$, $Z_B = \sum_{i\in B} Z_i$, and $Z_{AB} = \sum_{i\in A, j\in B} c_{ij}$, where Z_A is the partition function of the region A, Z_B is the partition function of the region B, and Z_{AB} is the partition function of the cutting surface (i.e., of the barrier) that divides the cFEP into A and B. Thus, the free-energy of the barrier can be written as $\Delta G = -kTln(Z_{AB})$. It is possible to isolate all the basins and barriers by iterative determinations of the minimum-cuts between all pair of nodes. In practice, to calculate the cFEP using the mfpt as progress coordinate, the nodes are sorted according to their mfpt value. For any mfpt_c between 0 and mfpt_{MAX} a point $[Z_A/Z, -kTln(Z_{AB}/Z)]$ on the cFEP can be calculated, where A is the set of all nodes with $\text{mfpt}_i < \text{mfpt}_c$ and B the set of nodes with $\text{mfpt}_i > \text{mfpt}_c$ (a schematic illustration is presented in Supp. Mat., Fig. S2).

The combination of the cFEP with a RMSD coarse-graining was particularly effective to

capture the difference between R- and L-configurations of Arg10. Compared with secondary structure coarse-graining used in previous works[6,7], RMSD coarse-graining has the main advantage that it takes into account side chain conformations. This feature allows us to study the R- to L-configuration transition mechanism. Furthermore, it does not group together conformations that are far from each other in configuration space (i.e., distant in terms of RMSD) simply because they have the same secondary structure string. RMSD clusters were defined as follows: firstly, the conformations sampled by molecular dynamics were superimposed considering only the C_α atoms and secondly, side chain atoms were used to group superimposed conformations into clusters having a side chain RMSD value below a chosen threshold. The C_α-atoms superimposition is justified because we are considering the helical ensemble (i.e., composed of similar structures without bending in the chain). It reduces the number of atoms needed in the latter part of the calculation, reducing the computational time (RMSD is a pairwise measurement). At the end, secondary structure strings are used to "label" the obtained clusters.

V. ACKNOWLEDGMENTS

The MD simulations were performed on the Matterhorn cluster of the University of Zurich and we gratefully acknowledge the support of C. Bolliger and A. Godknecht. This work was supported by a Swiss National Science Foundation grant to AC.

[1] C. A. Rohl and R. L. Baldwin, Meth. in Enzymol. **295**, 1 (1988).

[2] J. R. Kumita, O. S. Smart, and G. A. Woolley, Proc. Natl. Acad. Sci. USA **97**, 3803 (2000).

[3] F. Zhang, O. Sadovski, S. J. Xin, and G. A. Woolley, J. Am. Chem. Soc. **129**, 14154 (2007).

[4] J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, and P. Hamm, Proc. Natl. Acad. Sci. USA **102**, 2379 (2005).

[5] J. A. Ihalainen, J. Bredenbeck, R. Pfister, L. Chi, I. H. M. van Stokkum, G. A. Woolley, and P. Hamm, Proc. Natl. Acad. Sci. USA **104**, 5383 (2007).

[6] J. A. Ihalainen, B. Paoli, S. Muff, E. H. G. Backus, J. Bredenbeck, G. A. Woolley, A. Caflisch, and P. Hamm., Proc. Natl. Acad. Sci. USA **105**, 9588 (2008).

[7] B. Paoli, M. Seeber, E. Backus, J. Ihalainen, P.Hamm, and A. Caflisch, J. Phys. Chem. B (in press).

[8] H. S. Chan and K. A. Dill, Proteins: Structure, Function, and Bioinformatics **30**, 2 (1998).

[9] A. Caflisch, Curr. Opin. Struc. Biol. **16**, 71 (2006).

[10] S. V. Krivov and M. Karplus, J. Phys. Chem. B **110**, 12689 (2006).

[11] S. V. Krivov, S. Muff, A. Caflisch, and M. Karplus, J. Phys. Chem. B **112**, 8701 (2008).

[12] C. A. F. Andersen, A. G. Palmer, S. Brunak, and B. Rost, Structure **10**, 174 (2002).

[13] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[14] M. Seeber, M. Cecchini, F. Rao, G. Settanni, and A. Caflisch, Bioinformatics **23**, 2625 (2007).

[15] P. Ferrara, J. Apostolakis, and A. Caflisch, Proteins: Structure, Function, and Bioinformatics **46**, 24 (2002).

[16] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[17] F. Rao and A. Caflisch, J. Mol. Biol. **342**, 299 (2004).

[18] I. Hubner, E. Deeds, and E. Shakhnovich, Proc. Natl. Acad. Sci. USA **103**, 17747 (2006).

[19] S. V. Krivov and M. Karplus, Proc. Natl. Acad. Sci. USA **101**, 14766 (2004).

[20] J. Hartigan, Clustering Algorithms (John Wiley & Sons, Inc., New York, 1975).

[21] R. Koradi, M.Billeter, and K. Wüthrich, J. Mol. Graphics Modell. **14**, 51 (1996).

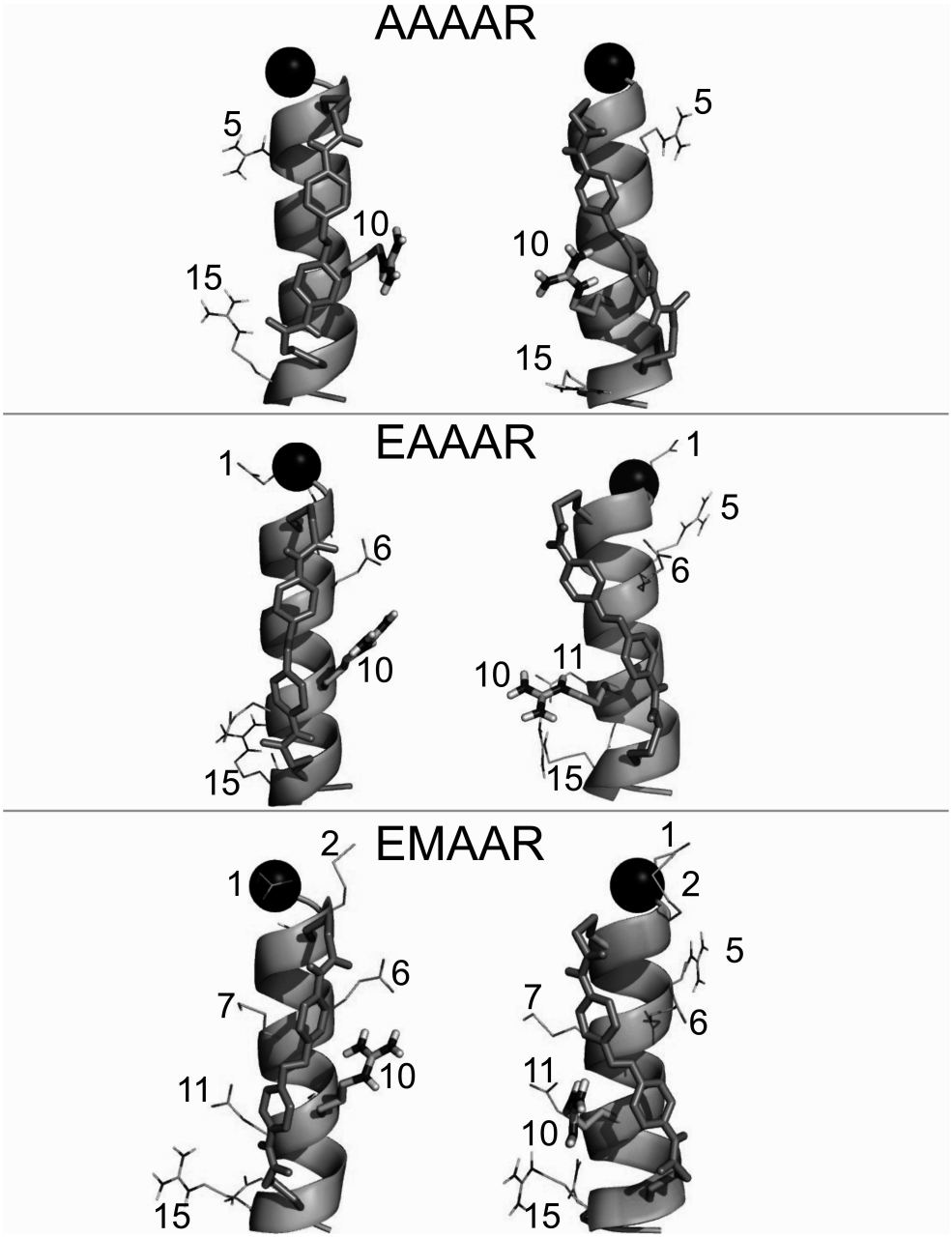


FIG. 1: The two orientations of Arg10 in the equilibrium trans simulations. Upon vertical alignment with N-terminus (black spheres) on top the R-configuration has Arg10 on the right side with respect to the cross-linker (left column) while the L-configuration has Arg10 on the left side (right column). The R- to L- transition (and viceversa) is hindered by the presence of the cross-linker.

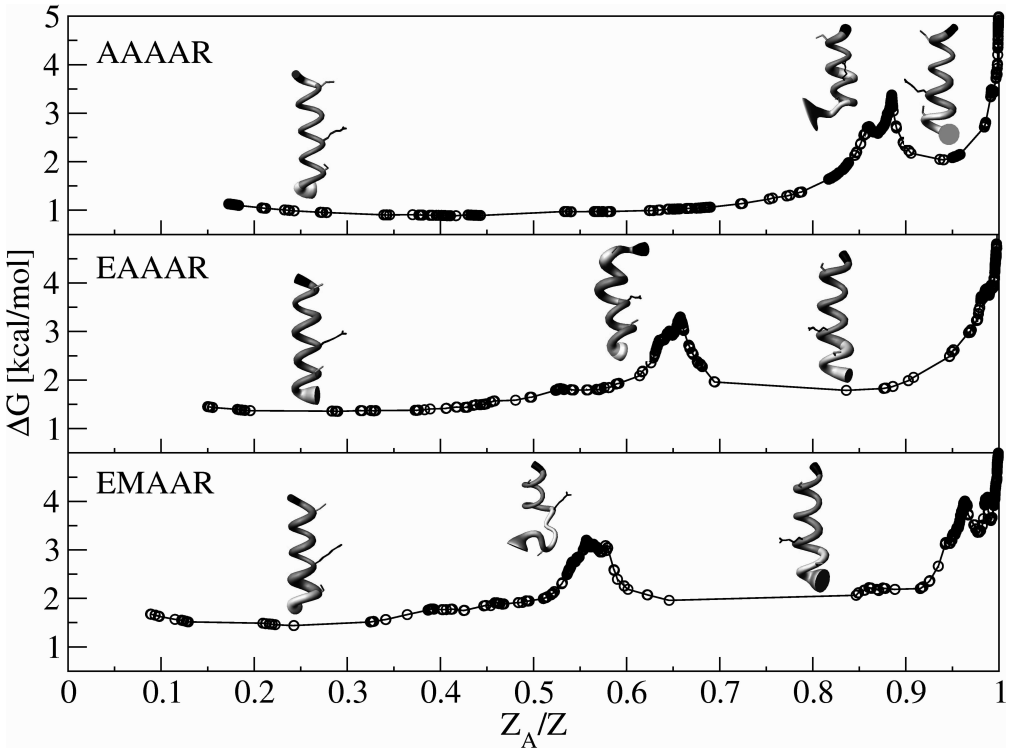


FIG. 2: The cFEP shows a free-energy barrier separating RMSD clusters with R-configuration of Arg10 from clusters with L-configuration. The height of the barrier is about 2 kcal/mol and it moves towards smaller values of the progress coordinate Z_A/Z on going from AAAAR to EAAAR to EMAAR. The representative structures, made with MOLMOL[21] are shown with flexible tubes of variable diameter reflecting the conformational disorder. The structures on the left side of the barriers have R-conformation of Arg10, whereas structures on the right side have L-configuration. The structures on the top of the barriers show the partial unfolding of the α-helix necessary to accomplish the R- to L- transition.

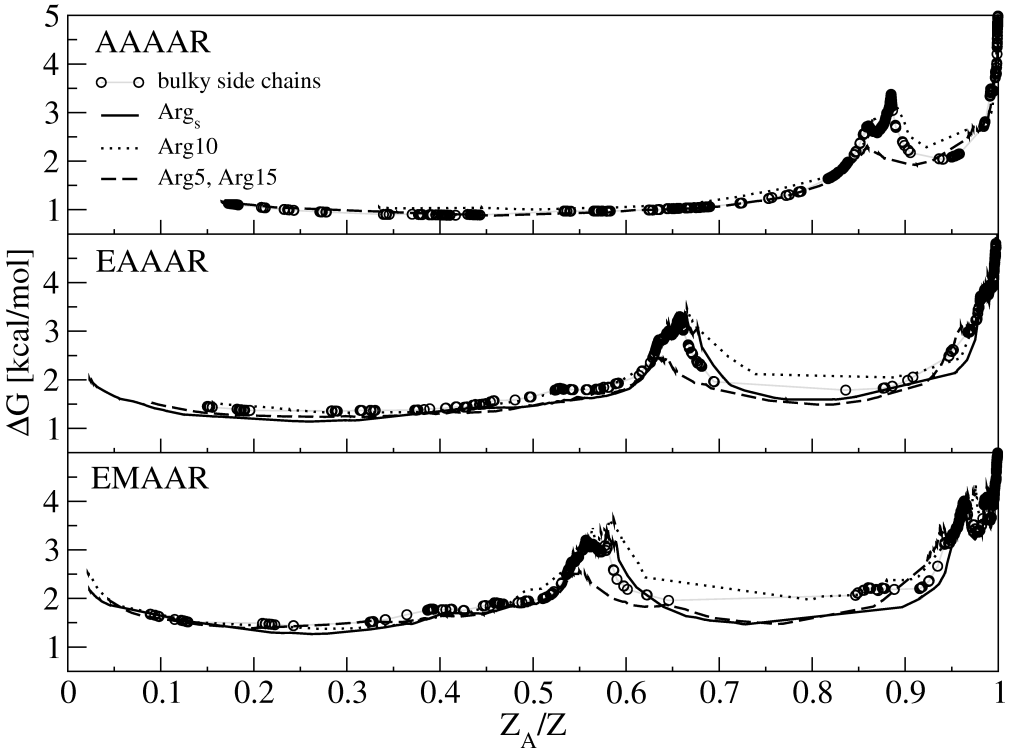


FIG. 3: Effect of different atom selections on the trans equilibrium FEPs at 281 K of AAAAR, EAAAR, and EMAAR. The main responsible of the barrier are the different configurations of Arg10: all the selections containing Arg10 give similar profiles with similar barrier heights. When Arg10 is removed from the selection (dashed curve), the barrier becomes much smaller, indicating that the effect on the FEPs of arginines different from Arg10 is much lower that the effect of Arg10.

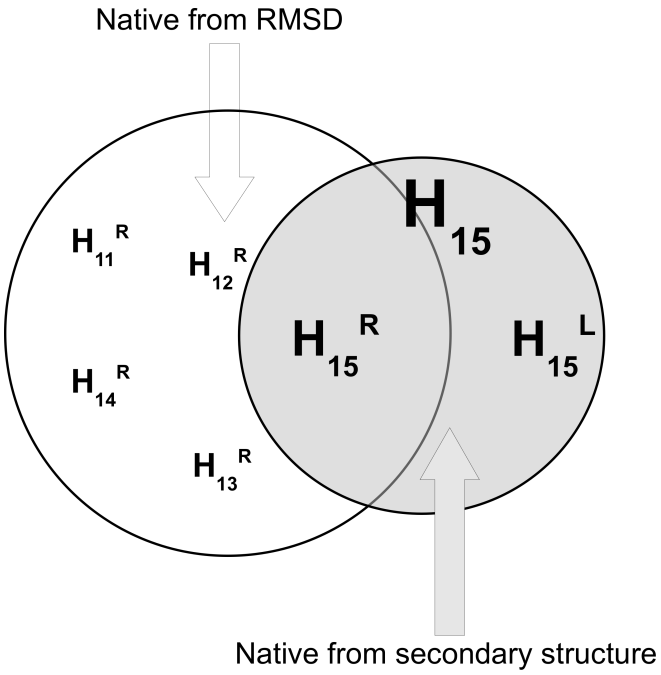
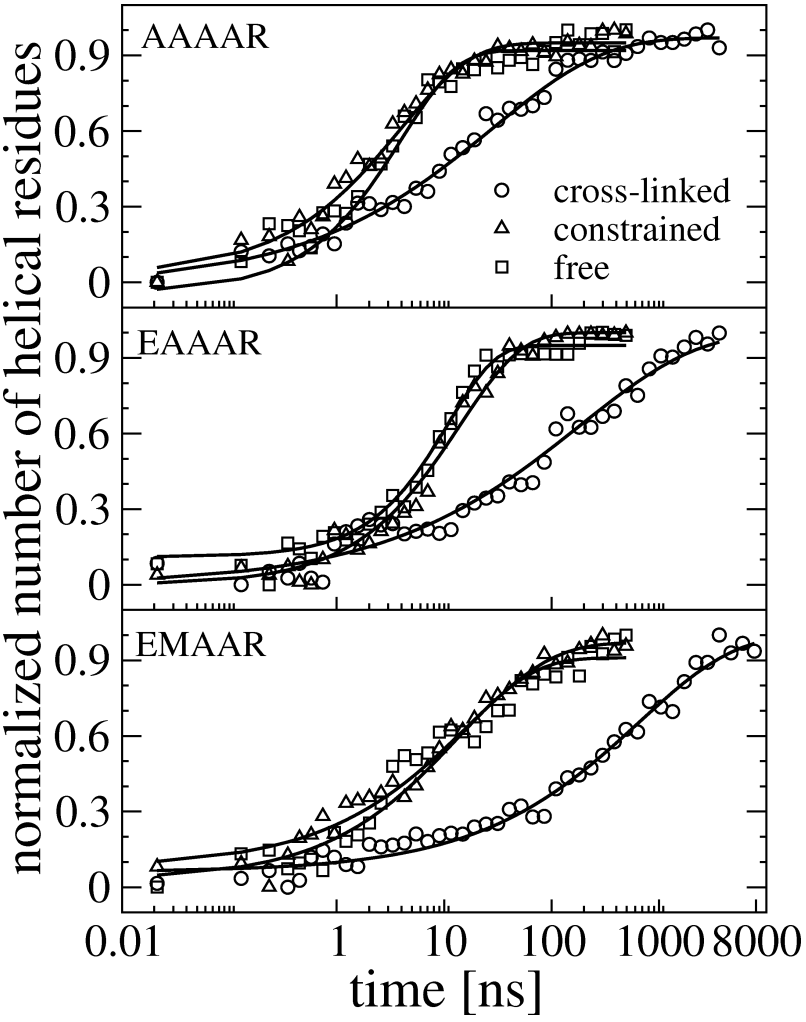


FIG. 4: Schematic illustration of the correct native state obtained with RMSD coarse-graining (white) and -H_{15}- string of secondary structure (light gray). The intersection represents the -H_{15}- string with R-configuration of Arg10. The secondary structure-based coarse-graining is not sufficient to determine the folded state because it does not take into account the orientation of the side chains.



	cross-linked		constrained		free	
	τ [ns]	β	τ [ns]	β	τ [ns]	β
AAAAR	28	0.42	4	0.66	4	0.83
EAAAR	185	0.40	14	0.76	11	1.00
EMAAR	653	0.49	17	0.55	13	0.61

FIG. 5: Normalized helical content (number of residues in α, π or 3_{10} conformation) along MD kinetic runs. MD data (symbols) are fitted with stretched-exponential functions (solid lines). The folding time increases with increasing number of long side chains and the increase is more pronounced for cross-linked peptides.

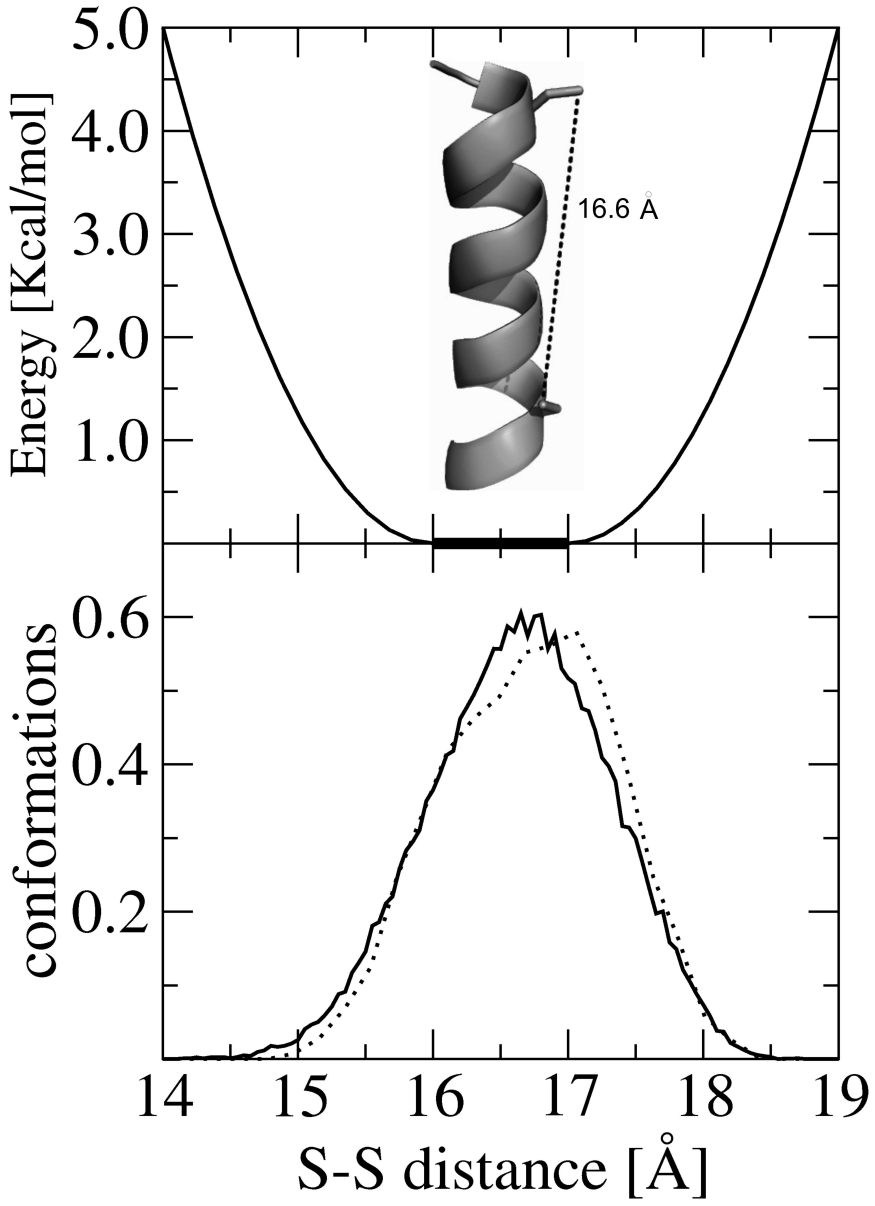


FIG. 6: Distance constraint used to emulate the cross-linker. (Top) The constraining function is harmonic at the borders and zero in the 16.0-17.0 Å interval. The average sulfur atom distance in the trans conformation of the cross-linker is 16.6 Å (inset). (Bottom) Sulfur atoms distance distribution when the linker is in trans conformation (solid curve). The dotted curve is the results obtained in a 100 ns kinetic run using a constraining function with R_{MAX}-R_{MIN}=1 Å and a force constant of 2.5 kcal/mol Å^2.

EQUILIBRIUM

	CIS		**TRANS**	
	Nr. simulations	length [μs]	Nr. simulations	length [μs]
AAAAR	2	24	4	12
EAAAR	1	18	2	11
EMAAR	3	27	2	16

TABLE I: Number of equilibrium simulations performed and corresponding time-length for cis and trans conformations of the cross-linker. The time-length values refer to individual replicas, and not to the cumulative time which is six times larger.

FOLDING RUNS

	cross-linked		**constrained**		**free**	
	Nr. of runs	length [μs]	Nr. of runs	length [μs]	Nr. of runs	length [μs]
AAAAR	100	4	50	0.5	50	0.5
EAAAR	50	4	50	0.5	50	0.5
EMAAR	100	8	50	0.5	50	0.5

TABLE II: Number of kinetic runs performed and corresponding time-length for cross-linked, constrained and free systems.

Cross-linked helical peptides have complex free energy surface

SUPPLEMENTARY MATERIAL

B. Paoli et al.

Department of Biochemistry,

University of Zurich,

Winterthurerstrasse 190,

CH-8057 Zurich, Switzerland

tel: +41 44 635 55 21,

fax: +41 44 635 68 62,

Keywords: molecular dynamics, distance constraint, stretched-exponential kinetics, free-energy profiles

I. SUBSTITUTION OF THE CROSS-LINKER WITH A CONSTRAINT

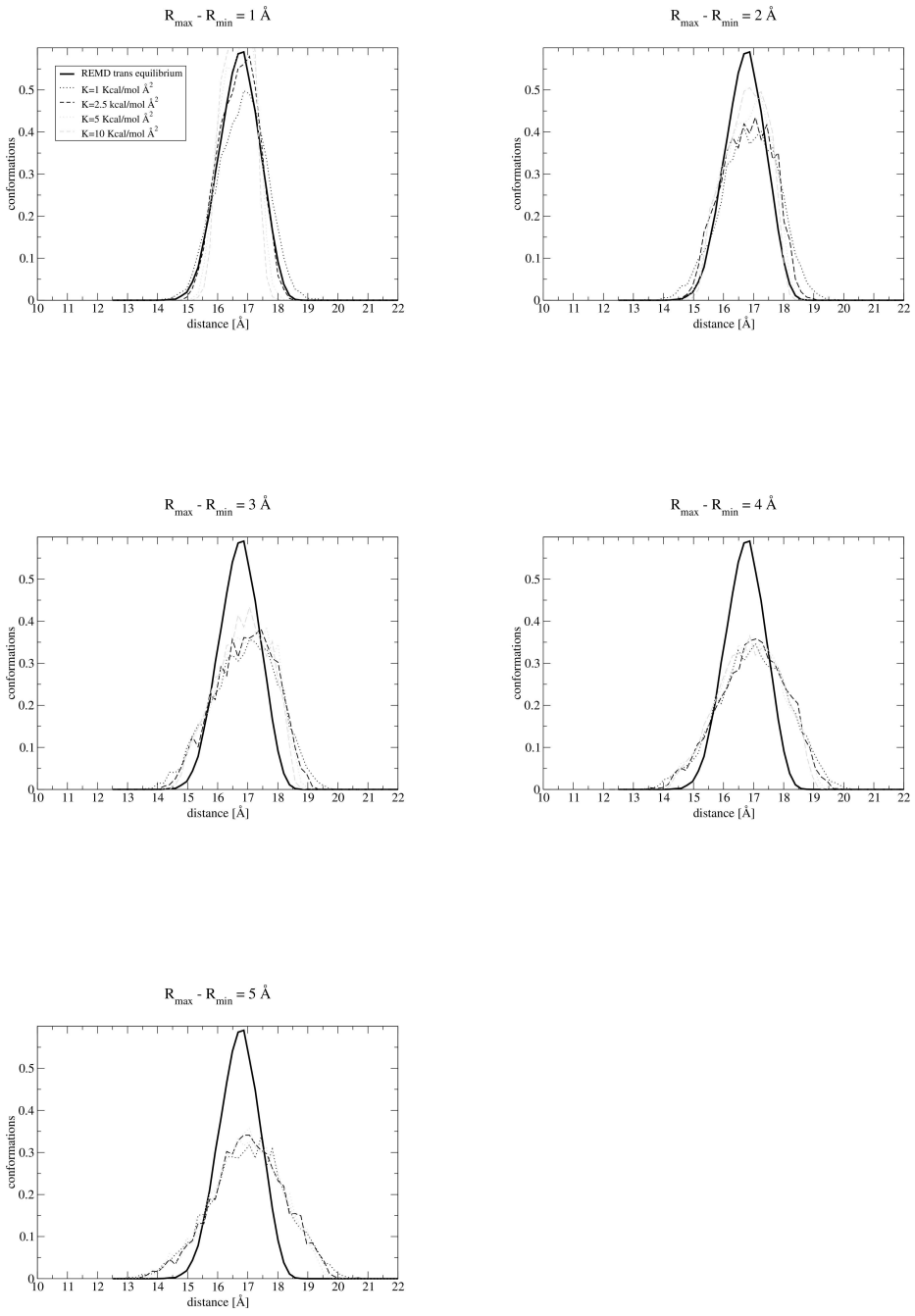


FIG. S1: Cysteine sulfur atoms distance distribution in a 100 ns kinetic test-run with R_{MAX}-R_{MIN} ranging from 1 to 5 Å and K ranging from 1 to 10 Kcal/mol Å^2, see also Tab. S-I

R_{min} [Å]	R_{max} [Å]	(R_{min}-R_{max}) [Å]
K=1 [Kcal/mol Å^2]		
14.0	19.0	5
14.5	18.5	4
15.0	18.0	3
15.5	17.5	2
16.0	17.0	1
K=2.5 [Kcal/mol Å^2]		
14.0	19.0	5
14.5	18.5	4
15.0	18.0	3
15.5	17.5	2
16.0	17.0	1
K=5 [Kcal/mol Å^2]		
14.0	19.0	5
14.5	18.5	4
15.0	18.0	3
15.5	17.5	2
16.0	17.0	1
K=10 [Kcal/mol Å^2]		
14.0	19.0	5
14.5	18.5	4
15.0	18.0	3
15.5	17.5	2
16.0	17.0	1

TABLE S-I: Summary of tests performed. R_{MAX}-R_{MIN} varies between 1 Å and 5 Å and K between 1 Kcal/mol Å^2 and 10 Kcal/mol Å^2.

II. MFPT-BASED CFEP

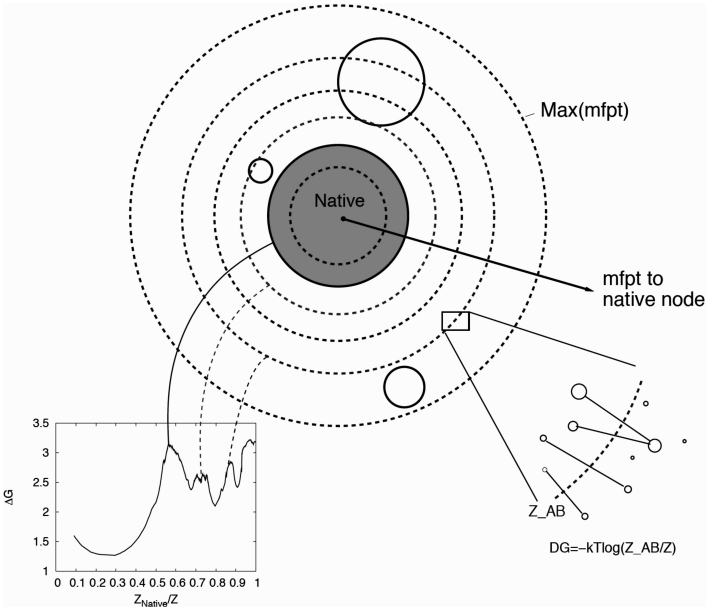


FIG. S2: Schematic illustration of the one-dimensional FEP procedure using mfpt as progress variable. Each of the four solid circles represents a free-energy basin while concentric dashed circles represent values of mfpt. For each value of mfpt_c between 0 (native node) and mfpt_{MAX} a point in the profile is obtained. Bottom right: ΔG of the fraction of links crossing the cutting surface at mfpt=mfpt_c. Bottom left: Relative partition function Z_A/Z, where the set A contains nodes with mfpt<mfpt_c. The figure is taken from Ref.[1]

A. Diffusivity

Diffusive regime is implicit for Monte Carlo simulations and valid in many cases in MD, as in protein folding studies. In this case the quenching interval dt should be small enough that the recrossings of the transition state (i.e. the barrier of the cFEP in Fig. S3, black curve) is taken into account. On the other extreme, we have the ballistic regime obtained when the quenching interval is large enough that the number of recrossings is negligible. To check whether our quenching interval is small enough (i.e. we are in the diffusive regime), we considered the dependence of the cFEPs on the quench interval dt [2]. The original dt is the chosen saving frequency for MD simulations, dt_1=20 ps. The one for the check is dt_2=40ps. If the profiles are consistent with the diffusive regime, they are proportional to $dt^{-\frac{1}{2}}$. In other words, $dt_2 = dt_1 + \frac{ln(dt_2)}{2}$. Since dt_2=2, we have that the condition for diffusive regime is $dt_2 = dt_1 + 0.35$. In Fig. S3, we compared the FEPs for AAAAR, EAAAR, and EMAAR calculated for dt_1 (black curves) and dt_2 (red curves). The evaluation of the diffusivity $dt_2 = dt_1 + 0.35$ (dashed blue curves) is good.

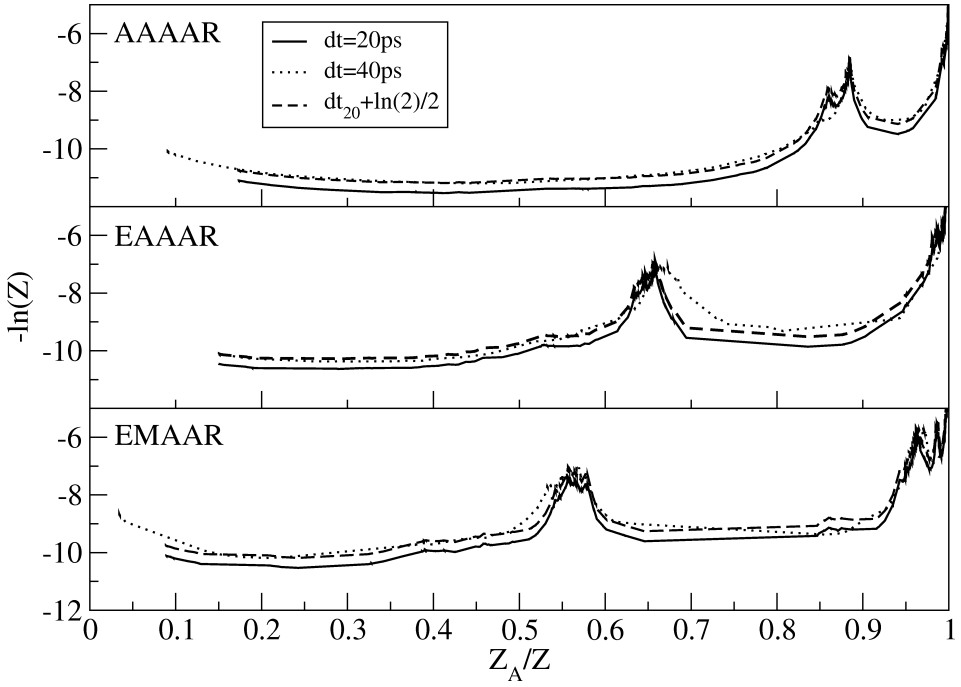


FIG. S3: Diffusivity check for AAAAR, EAAAR, and EMAAR peptides. The solid curves are calculated for dt=20 ps, the dotted curves for dt=40 ps, and the dashed curves are calculated as $dt_2 = dt_1 + ln(2)/2$

III. MOST POPULATED RMSD CLUSTER WITH R-CONFIGURATION OF ARG10

	pop./cluster [%]	pop./total [%]
AAAAR		
-HHHHHHHHHHHHHH-	59	10
-HHHHHHHHHHHHHI-	18	3
-HHHHHHHHHHHHHT-	11	2
-HHHHHHHHHHHHII-	6	1
-HHHHHHHHHHHHH--	2	0.4
-IIHHHHHHHHHHHH-	0.8	0.1
--HHHHHHHHHHHH-	0.6	0.1
-HHHHHHHHHHHHG-	0.4	0.06
-HHHHHHHHHHHIII-	0.4	0.06
-IHHHHHHHHHHHH-	0.3	0.04
total	**98.5**	**16.8**
EAAAR		
-HHHHHHHHHHHHHH-	31	1
-HHHHHHHHHHHHHI-	22	1
-HHHHHHHHHHHHHT-	15	0.6
-HHHHHHHHHHHHII-	15	0.3
-HHHHHHHHHHHHH--	3	0.1
-IIHHHHHHHHHHHH-	1	0.06
-HHHHHHHHHHHHG-	1	0.05
--HHHHHHHHHHHH-	0.7	0.03
-GHHHHHHHHHHHH-	0.6	0.03
-THHHHHHHHHHHH-	0.6	0.02
total	**97.8**	**14.8**
EMAAR		
-HHHHHHHHHHHHII-	31	1
-HHHHHHHHHHHHH-	22	1
-HHHHHHHHHHHIII-	15	0.6
-HHHHHHHHHHHHI-	15	0.3
--HHHHHHHHHHHH-	3	0.1
-HHHHHHHHHHHHT-	1	0.06
-HHHHHHHHHHIIII-	1	0.05
-HHHHHHHHHHHI--	0.7	0.03
-HHHHHHHHHHGGG-	0.6	0.03
-HHHHHHHHHIIIII-	0.6	0.02
total	**89.9**	**3.2**

TABLE S-II: Ten most populated secondary structure strings in the heaviest side chains RMSD cluster with R-configuration of Arg10 for AAAAR, EAAAR and EMAAR peptides forthe trans equilibrium simulations at 281 K. The most populated conformation of the heaviest cluster is labeled, in terms of secondary structure strings, as $\texttt{-H}_{15}\texttt{-}$ and shows R-configuration for AAAAR and EAAAR peptides and L-configuration for EMAAR peptide. For EMAAR the heaviest cluster with R-configuration was used as target for cFEP and reported here. Note that in EMAAR peptide the string labeled as $\texttt{-H}_{15}\texttt{-}$ is the second one with a population of 22% and the population of strings having some π-helix (I in the DSSP code[3]) is higher compared to other peptides. This is because the insertion of methionines at positions n, n+5 along the sequence may favor the formation of π-helix as reported by Hiltpold et al.[4].

IV. MOST POPULATED RMSD CLUSTER WITH L-CONFIGURATION OF ARG10

	pop./cluster [%]	pop./total [%]
AAAAR		
-HHHHHHHHHHHHHH-	26	0.8
-HHHHHHHHHHHHIII-	23	0.7
-HHHHHHHHHHHIIII-	18	0.6
-HHHHHHHHHHHHHI-	9	0.3
-HHHHHHHHHHIIIII-	7	0.2
-HHHHHHHHHHHHII-	5	0.2
-HHHHHHHHHHHII--	2	0.1
-HHHHHHHHHHIII--	1	0.05
-HHHHHHHHHHHGGG-	0.9	0.03
-HHHHHHHHHIHHHH-	0.9	0.03
total	**92.8**	**3.0**
EAAAR		
-HHHHHHHHHHHHHH-	33	5
-HHHHHHHHHHHIII-	22	3
-HHHHHHHHHHIIII-	17	2
-HHHHHHHHHHHHHI-	12	2
-HHHHHHHHHIIIII-	5	0.7
-HHHHHHHHHHHHII-	4	0.6
-HHHHHHHHHHHII--	2	0.2
-HHHHHHHHHIHHHH-	1	0.1
-HHHHHHHHHHIII--	0.6	0.08
-HHHHHHHHHHHGGG-	0.5	0.07
total	**97.1**	**13.8**
EMAAR		
-HHHHHHHHHHHHHH-	29	3
-HHHHHHHHHHHIII-	16	2
-HHHHHHHHHHHHHI-	14	1
-HHHHHHHHHHIIII-	14	1
-HHHHHHHHHHHHII-	6	1
-HHHHHHHHHIIIII-	6	0.6
--HHHHHHHHHHHHH-	2	0.2
-HHHHHHHHHIHHHH-	1	0.1
-IIHHHHHHHHHHHH-	0.8	0.08
--HHHHHHHHHHHHI-	0.6	0.06
total	**89.4**	**9.0**

TABLE S-III: Same as Table S-II for the ten most populated secondary structure strings of the heaviest cluster with L-configuration of Arg10. The most populated secondary structure string is again $-H_{15}-$ though with a lower weight.

[1] S. V. Krivov, S. Muff, A. Caflisch, and M. Karplus, J. Phys. Chem. B **112**, 8701 (2008).

[2] S. Krivov and M. Karplus, Proc. Natl. Acad. Sci. USA **105**, 13841 (2008).

[3] C. A. F. Andersen, A. G. Palmer, S. Brunak, and B. Rost, Structure **10**, 174 (2002).

[4] A. Hiltpold, P. Ferrara, J. Gsponer, and A. Caflisch, J. Phys. Chem. B **104**, 10080 (2000).

Conclusions and Outlook

The protein folding problem is one of the most difficult challenge of the modern experimental and theoretical biophysics. The question arises on how proteins fold to their native structure, the pre-requisite to fulfil their function. This means not only to find out the relation between the amino acid sequence and the three dimensional structure, but also to explain the hidden mechanisms through which a protein reaches its lowest free energy configuration. The problem is intrinsically complex due to the large number of degrees of freedom involved [1]. Much of the complexity found in protein folding can also be observed in the dynamics of smaller polypeptides that undergo secondary structure formation [2–5]. Such small peptides are the subject of the present dissertation. There are two main reasons to study small peptides. First, they mimic protein in complexity but, computationally speaking, are small enough to allow detailed simulation studies (especially when implicit solvent models are employed) [6]. Second, the development of fast (nanosecond) time-resolved spectroscopy methods allows to study peptide folding dynamics on the same timescale as computer simulations [7–12]. Moreover, peptides consisting of a single α-helix give the opportunity to compare theory and experiments because their small size allows extensive sampling by simulations. In particular, the employment of α-helical peptides with enhanced

helix propensity have been used to study the physico-chemical properties of an α-helix [13]. Additionally, the incorporation of a photoswitchable cross-linker, has been established as an efficient experimental tool to control the helix stability and to reversibly control secondary structure content [14].

The work presented in this thesis is devoted to the study, by means of molecular dynamics (MD) simulations of the folding of three model cross-linked (i.e., structurally constrained) peptides having different amount of bulky side chains. The significant changes of the folding kinetics observed as a function of temperature in both time resolved infrared experiments [15, 16] and MD simulations [17, 18] are indicative of the complexity of the folding process. To shed light onto the complex behavior, two approaches were employed: Kinetic Grouping Analysis (KGA) and cut-based free-energy profiles (cFEP). Both these methods group conformations not according to structural similarity criterion but rather according the transitions observed during the simulations, or *kinetic similarity*. The KGA method was successful in the identification of metastable states from MD folding runs of the cross-linked α-helical peptide [17,18]. The cFEP suggested that the native state of cross-linked peptides, as calculated using secondary structure based coarse-graining, is not described accurately because it does not take into account the side chains orientation with respect to the cross-linker [19]. To investigate the role of the entanglement of bulky side chains with the cross-linker, we substituted it with a distance constraint mimicking the effect of a non-bulky cross-linker. The possibility of decoupling the twofold effects of the cross-linker on the backbone and the side chains, possible only in the context of MD simulations, suggested that the main effect of the cross-linker on the overall kinetics is due to its entanglement with the side-chains rather that its effect as a distance constraining.

The insights achieved through MD simulations were shown to be consistent with the experimental studies at least at a qualitative level. Thus the present work clarified questions which would not have been easily answered by experiments only.

Finally, one important issue concerning our MD simulations is the possible presence of small biases due to the implicit solvation model [20]. Although its reliability is widely accepted, it was also pointed out in the past that it may tend to favor helix formation [21]. This should represent only a minor effect in our case, since all the studied peptides are intrinsically characterized by a strong helix propensity. It is nevertheless important to check, as a long-term aim, this issue by carrying out simulations in explicit water.

Bibliography

[1] M. Karplus. Aspects of protein reaction dynamics: deviations from simple behavior. *J. Phys. Chem. B*, 104:11–27, 2000.

[2] V. Muñoz, P.A. Thompson, J. Hofrichter, and W.A. Eaton. Folding dynamics and mechanism of β-hairpin formation. *Nature*, 390:196–199, 1997.

[3] S. Gnanakaran, H. Nymeyer, J. Portman, K.Y. Sanbonmatsu, and A.E. Garcia. Peptide folding simulations. *Curr. Opin. Struct. Biol.*, 13:168–174, 2003.

[4] R.B. Dyer, S.J. Maness, E.S. Peterson, S. Franzen, R M. Fesinmeyer, and N.H. Andersen. The mechanism of β-hairpin formation. *Biochemistry*, 43:11560–11566, 2004.

[5] S.V. Krivov and M. Karplus. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. USA*, 101:14766–14770, 2004.

[6] A. Hiltpold, P. Ferrara, J. Gsponer, and A. Caflisch. Free energy surface of the helical peptide Y$(\text{MEARA})_6$. *J. Phys. Chem. B*, 104:10080–10086, 2000.

[7] S. Williams, T.P. Causgrove, R. Gilmanshin, K.S. Fang, R.H. Callender, W.H. Woodruff, and R.B. Dyer. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, 35:691–697, 1996.

[8] P.A. Thompson, W.A. Eaton, and J. Hofrichter. Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a "kinetic zipper" model. *Biochemistry*, 36:9200–9210, 1997.

[9] M. Gruebele, J. Sabelko, R. Ballew, and J. Ervin. Laser Temperature Jump Induced Protein Refolding. *Acc. Chem. Res.*, 31:699–707, 1998.

[10] C.Y. Huang, J.W. Klemke, Z. Getahun, W.F. DeGrado, and F. Gai. Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc.*, 123:9235–9238, 2001.

[11] J. Hofrichter. Laser temperature-jump methods for studying folding dynamics. *Methods Mol. Biol.*, 168:159–191, 2001.

[12] R. Callender and R.B. Dyer. Probing protein dynamics using temperature jump relaxation spectroscopy. *Curr. Opin. Struc. Biol.*, 12:628–633, 2002.

[13] C.A. Rohl and R.L. Baldwin. Deciphering rules of helix stability in peptides. *Meth. in Enzymol.*, 295:1–26, 1998.

[14] J.R. Kumita, O.S. Smart, and G.A. Woolley. Photo-control of helix content in a short peptide. *Proc. Natl. Acad. Sci. USA*, 97:3803–3808, 2000.

[15] J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, and P. Hamm. α-Helix formation in a photoswitchable peptide tracked from picosec-

onds to microseconds by time-resolved IR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 102:2379–2384, 2005.

[16] J. A. Ihalainen, J. Bredenbeck, R. Pfister, L. Chi, I. H. M. van Stokkum, G. A. Woolley, and P. Hamm. Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc. Natl. Acad. Sci. USA*, 104:5383–5388, 2007.

[17] J. A. Ihalainen, B. Paoli, S. Muff, E. H. G. Backus, J. Bredenbeck, G. A. Woolley, A. Caflisch, and P. Hamm. α-Helix folding in the presence of structural constraints. *Proc. Natl. Acad. Sci. USA*, 105:9588–9593, 2008.

[18] B. Paoli, M. Seeber, E.H.G. Backus, J.A. Ihalainen, P.Hamm, and A. Caflisch. Bulky side chains and non-native salt bridges slow down the folding of a cross-linked helical peptide: A combined molecular dynamics and time-resolved infrared spectroscopy study. *J. Phys. Chem. B, in press*, 2009.

[19] B.Paoli et al. Cross-linked helical peptides have complex free-energy surface OR Hidden complexity of free-energy surfaces of cross-linked helical peptides. *to be submitted*, 2009.

[20] P. Ferrara, J. Apostolakis, and A. Caflisch. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 46:24–33, 2002.

[21] P. Ferrara, J. Apostolakis, and A. Caflisch. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B*, 104:5000–5010, 2000.

Curriculum Vitae

Beatrice Paoli

born 13^{th} April 1981 in Rome, Italy

Italian citizen

Education

March 2009	Successfull doctoral examination (Dr. Sc. Nat.)
October 2005–March 2009	PhD student in the group of Prof. Dr. A. Caflisch at the Department of Biochemistry, University of Zurich
February 2005–July2005	Postdiploma research at the University of Rome "La Sapienza".
January 2005	Physics Degree (Italian "Laurea") at the University of Rome "La Sapienza". Score: 110/110 *cum laude*.
1999–2005	Studies of Physics at the University of Rome "La Sapienza"
1994–1999	High School Diploma at Liceo Scientifico Statale "C. Cavour" in Rome. Score 100/100.